Stony Brook University

Two Essays in Quantitative Finance

A Dissertation presented

by

Ke Wang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

August 2016

Stony Brook University

The Graduate School

Ke Wang

We, the dissertation committe for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Haipeng Xing - Dissertation Advisor
Associate Professor, Applied Mathematics and Statistics,SUNYSB

Svetlozar Rachev
Professor of Finance, College of Business, SUNYSB

Raphael Douady
Professor, QF Frey Endowed Chair, Applied Mathematics and Statistics,SUNYSB

Keli Xiao
Assistant Professor of Finance, College of Business, SUNYSB

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Two Essays in Quantitative Finance

by

Ke Wang

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2016

This dissertation explores two interesting problems in quantitative finance. In the first part, we consider the detection methods on structural breaks that are characterized by a credit rating transition matrix based on homogeneous Markov process model. Recent studies have shown that firms' credit rating migration process is not stationary and may have structural breaks. Assuming the generator of probability transition matrices of firms' credit rating to be piecewise constant and the jump time of generator corresponds to the structural break time in the pattern of firms' rating migrations, we study several types of sequential surveillance rules for early detection. The surveillance rules we investigated includes the Shewhart control chart, an generalized likelihood ratio (GLR) detection rule for a single change-point

with unknown pre- and post-change transition matrices, a detection rule based on an extension of Shiryaev's Bayes single change-point model, and a detection rule for multiple unknown structural breaks. We provide theoretical discussion and extensive simulations to compare the performance of these rules. We further use these rules to online detect structural breaks in firms' credit rating migrations based on U.S. firms' rating record from 1986 to 2012. In part two, we develop a multivariate log-linear Poisson time series model to investigate the interdependence between components of a vector time series of counts. Maximum likelihood method is used for the estimation of the parameters and the property of geometrically ergodic is demonstrated. We further successfully applied it to study the interdependence of trading behavior in high-frequency trading market and of tail exceedance events in different markets. Specifically, we generalized the univariate log-linear Poisson model for time series counts data to the multivariate case, and developed an inference procedure for it. In this study, this model has been applied to investigate two types of time series counts data in finance. The first application is to use the developed model to study the dependence of financial risks in different market. Specifically, consider the stock market indices in the US, Europe, and Japan, the exceedance of the stock index return over certain threshold represents the magnitude of market variations and provides us a new measurement for the market tail risk in different countries/market. The second application is the interdependence of trading behavior for different stocks, through which the impact of one stock's trading behavior on another stock can be quantitatively modeled and identified by this model.

To my parents

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Haipeng Xing, for exciting topics suggested, his excellent guidance, continuous support, caring and patience during my thesis research. Indeed, this dissertation would never have been completed without his enthusiasm, dedication and great support.

I would also like to thank Professor Svetlozar Rachev, Professor Raphael Douady and Professor Keli Xiao. It is my great honor to have them on my dissertation committee.

I would like to thank Yan Liu for his support, encouragement, patience and love in the past twelve years and my parents Mingwei Wang and Liping Xu, for their unconditional support and encouragement in all my educational endeavors. Without their encouragement, I would never have been able to finish my Ph.D. studies and without their love, all my achievement would mean nothing for me.

I would also like to thank my family, especially my aunt Shuyuan Xu, my uncle Shixue Sun, both of whom have passed away but always cared my study.

Last but not least, I would like to thank all my groupmates and friends, especially Yang Yu, Xiaojin Dong and Ying Cai, for all their kindness, friendship, and help.

Part I

Surveillance of Structural Breaks in Credit Rating Dynamics

Chapter 1

Introduction

1.1 Literature review

In modern credit risk management, people usually assume that the credit rating of firms follows a time homogeneous Markov Chain which is characterized by a credit rating transition matrix (Xing et al.,2012). There are some applications on using homogeneous Markov process to model credit rating transition matrix which summarizes historical data and changes of obligors' credit ratings in finance. For example, the measurement of risk for loans and bonds in the portfolio risk assessment is based on the joint distribution of rating transitions (Das et al., 2006; Frey and McNeil, 2007; Egloff et al., 2007; Duffie et al., 2009; Tsaig et al., 2011). In the

pricing model of bond and credit derivatives, the credit rating of obligors is used to value the risky credit derivatives (Jarrow et al., 1997, 1998; Lando, 2000; Acharya et al., 2006). In the credit rating industry, the credit rating transition reports are studied by the estimation of rating transition matrices.

Usually, people use a discrete-time setting to estimate the credit rating transition matrices. Now, the estimation of credit rating transitions based on continuous-time homogeneous Markov process is assumed for the rating process due to the rich significant dependence on regressors and the availability of rating data, also the advantages of continuous time Markov approach used instead of discrete one (Lando and Skødeberg, 2002; Bangia et al., 2002; Frydman and Schuerman, 2008). Based on Xing et al.(2012), particularly, suppose there are K rating classes where the state K is an absorbing state (e.g. bankruptcy). Also the rating transition process in the period $(0,t)$ is determined by the transition matrix $P(0,t)$, which is a continuous time homogeneous Markov Chain. The entry of ij represents the transition probability from class i to class j during the period $(0,t)$. The matrix $P(0,t)$ is represented using a generator matrix Λ, for $t>0$,

$$P(0,t) = exp(\Lambda t) := \sum_{k=0}^{\infty} \frac{\Lambda^k t^k}{k!} \tag{1.1.1}$$

where $\Lambda = (\lambda^{(i,j)})$ satisfies $\lambda^{(i,i)} = -\sum_{i\neq j} \lambda^{(i,j)}$ for $1 \leq i \leq K$, and $\lambda^{(i,j)} \geq 0$ for $1 \leq i \neq j \leq K$. Maximum likelihood estimation is used to estimate Λ to get $\widehat{\lambda}^{(i,j)} = \widehat{N}_{ij} / \int_0^t Y_i(s)ds$, in which $\widehat{N}_{ij}$ is the total number of transitions from i to

$j(\neq i)$ in $(0,t)$ and $Y_i(s)$ is the number of firms in rating class i at time s. (Küchler and Sørensen, 1997).

However, some publications have challenged the assumptions of time homogeneous Markov process of rating process. In Altman (1998)'s study, the credit rating migration is based on an in-depth investigation of the expected ratings changes (drift). Nickell et al. (2000)'s analysis of the stability of rating transitions considers the variations on time and the industry heterogeneity. Bangia et al. (2002) proposes that the volatility is a key part of a useful conceptual framework for stress testing credit portfolios. Credit ratings are sometimes considered stable over credit cycles because the market participants want ratings to be a view of an issuer's fundamental credit risk(Fons, 2002; Cantor and Mann, 2003; Altman and Rijken, 2004; Bruche et al., 2010). Frydman and Schuermann (2008) propose a mixture model of two independent continuous time homogeneous Markov chains for rating transitions process based on the assumptions of firms with same rating migration at different speeds without using a firm-specific information and analyze corporate credit rating history from Standard and Poor's spanning 1981-2002 basing on the proposed model. Weissbach and Walter (2010) study the time-stationarity of rating transitions that are modeled by a time-continuous discrete-state Markov process and derive a likelihood ratio test. They apply their approach to an internal rating data set, which reveals highly significant nonstationarity. Xing et al. (2012) model rating transition process as piecewise homogeneous Markov chains with unobserved structural breaks. Their proposed model provides explicit formulas for the posterior distribution of

the time-varying rating transition generator matrices, the probability of structural break at each period and prediction of transition matrices in the presence of possible structural breaks. Estimating the model by credit rating history, Xing et al. (2012) show that the structural breaks in rating transitions can be captured by the proposed model, and compare the prediction performance of their proposed and time homogeneous Markov chain models.

1.2 Motivation for structural breaks detection

Recently, testing on structural breaks has been an interesting research topic due to the important impact after it happens. Levine (2001) uses the test in Perron (1989)'s paper on structural breaks to evaluate changes of stock market liquidity after the policy change date. People has realized that the structural breaks has contributed to weakening risk management (Noel, 2008). Finding a better way to assess systemic risk has been an important area to reduce the risk of crises. The importance and advantages of detection on structural breaks has been more obvious. Valentinyi-Endrész (2004)'s study has given the result of whether detecting structural breaks in the volatility model can improve the Value at Risk forecast. Our study in this part is to develop statistical surveillance tools to monitor the stability of US firmsâĂŹ credit rating transition dynamics. An important and challenging problem after the 2007-2008 financial crisis is to monitor the stability of credit market and early detect the sharp changes of the market.

We should note that the model proposed by Xing et al. (2012) essentially identify historical market structural breaks or sharp changes based only on credit rating records. Basing on Xing et al. (2012), we assume the rating transition process of an obligor follows a K-state non-homogeneous continuous time Markov process, which is characterized by a transition probability matrix $P(s,t)$ in the period (s,t). The ijth element of $P(s,t)$ represents the probability of the obligor's rating being i at time s, while j at time t. Suppose there are n rating transitions observed in (s,t). For a transition time t_i in (s,t), denote $\Delta N_{kj}(t_i)$ the number of transitions observed from state k to stake j at time t_i, $\Delta N_k(t_i) = \sum_{1\le j\le K, j\neq k} \Delta N_{kj}(t_i)$, and $Y_k(t_i)$ the number of firms in state k right before time t_i. The transition matrix $P(s,t)$ can be consistently estimated by the product-limit estimator

$$\hat{P}(s,t) = \prod_{i=1}^{n} (I + \Delta \hat{A}(t_i)),$$

in which

$$\Delta \hat{A}(t_i) = \begin{pmatrix} -\frac{\Delta N_1(t_i)}{Y_1(t_i)} & \frac{\Delta N_{12}(t_i)}{Y_1(t_i)} & \frac{\Delta N_{13}(t_i)}{Y_1(t_i)} & \cdots & \frac{\Delta N_{1K}(t_i)}{Y_1(t_i)} \\ \frac{\Delta N_{21}(t_i)}{Y_2(t_i)} & -\frac{\Delta N_2(t_i)}{Y_2(t_i)} & \frac{\Delta N_{23}(t_i)}{Y_2(t_i)} & \cdots & \frac{\Delta N_{2K}(t_i)}{Y_2(t_i)} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \frac{\Delta N_{K-1,1}(t_i)}{Y_{K-1}(t_i)} & \frac{\Delta N_{K-1,2}(t_i)}{Y_{K-1}(t_i)} & \cdots & -\frac{\Delta N_{K-1}(t_i)}{Y_{K-1}(t_i)} & \frac{\Delta N_{K-1,K}(t_i)}{Y_{K-1}(t_i)} \\ 0 & 0 & \ldots & \ldots & 0 \end{pmatrix}$$

see Andersen et al. (1995). Here, the element of the kth diagonal represents the proportion of the firms $Y_k(t_i)$ leaving the state at time t_i, and the element of the

kjth off-diagonal represents the proportion of transitions from state k to state j at time t_i. Also, the kth state which is zero represents absorbing (i.e., default state). In the study of Xing et al. (2012), the non-homogeneous continuous time Markov process is decomposed into piecewise homogeneous continuous time Markov process with unobserved structural breaks.

1.3 Outline

To make an effort in surveillance of structural breaks as early as possible after it happens, this report in this part studies the detection of structural breaks in credit ratings based on the model proposed in Xing et al. (2012), using different detection rules on single change point and multiple change points. We have studied several types of sequential surveillance rules for early detection. The surveillance rules in this study include the Shewhart control chart, an generalized likelihood ratio (GLR) detection rule for a single change-point with unknown pre- and post-change transition matrices, a detection rule based on an extension of Shiryaev's Bayes single change-point model, and a detection rule for multiple unknown structural breaks. This study also provides theoretical discussion and extensive simulations to compare the performance of these rules, and further applies these rules to online detect structural breaks in firms' credit rating migrations based on U.S. firms' rating record from 1986 to 2012.

Chapter 2

Stochastic Structural Break Model and Associated Detection Rules

2.1 Shewhart Control Chart Detection Method

Based on the statistical surveillance of the CIR Model, proposed by Schmid and Tzothchev (2004), we modify the decision statistic $\mathbf{Z}_t$ based on an EWMA recursion on singular value decomposition (SVD) $M(P)$ as

$$\mathbf{Z}_t = (1-\alpha)\mathbf{Z}_{(t-1)} + \alpha M(P) \tag{2.1.1}$$

where the SVD metric for a transition matrix P is defined as

$$M(P) = \frac{1}{K}\sum_{i=1}^{K} \sqrt[2]{e_i(P-I)'(P-I)} \tag{2.1.2}$$

in which I is a $K \times K$ identity matrix and $e_i(\cdot)$ is the ith eigenvalue of the matrix.

Combining the above model, we derive a simple change-point detection rule, with a change-point at t, but not before t_p as

$$T = inf\{t > t_p : \mathbf{Z}_t \geq \theta\} \tag{2.1.3}$$

2.2 Generalized Likelihood Ratio (GLR) Detection Rule For a Single Change-Point Model

The observations, $\mathbf{Y}_t$ are independent random vectors with a common density function f_0 for $t < v$, and f_1 when $t \geq v$. Shiryaev (1978) gave an optimal sequential detection of the change-time v by Bayesian approach with following assumptions: (1) a loss of c for each observation taken at or after v; (2) a loss of 1 for a false alarm before v; (3) a geometric prior distribution on v. Based on the above assumptions, he proved that there would be an alarm as soon as the posterior probability exceeds some fixed level when a change has occurred using optimal stopping theory. Since

$$P\{v \leq n|\mathbf{Y}_1, \ldots, \mathbf{Y}_n\} = R_{p,n}/(R_{p,n} + p^{-1}) \tag{2.2.1}$$

where p is the parameter of the geometric distribution $P\{v = n\} = p(1-p)^{n-1}$ and $R_{p,n} = \sum_{k=1}^{n} \prod_{i=k}^{n} \{f_1(\mathbf{Y}_i)/(1-p)f_0(\mathbf{Y}_i)\}$, the Bayes rule declares at time

$$N(\gamma) = \inf\{n \geq 1 : R_{p,n} \geq \gamma\} \tag{2.2.2}$$

where a change has occurred. Roberts (1966) proposed the case of $p = 0$ in (2.2.2), and Pollak (1985) gave the Shiryaev-Roberts rule, which can be expressed as

$$\widetilde{N}(\gamma) = \inf\{n \geq 1 : \sum_{k=1}^{n} \prod_{i=k}^{n} (f_1(\mathbf{Y}_i)/f_0(\mathbf{Y}_i)) \geq \gamma\} \tag{2.2.3}$$

He also proved when $p \to 0$, it is asymptotically Bayes risk efficient.

Instead of Bayesian approach, Lorden (1971) minimized the worst-case expected delay using the minimax approach.

$$\overline{E}_1(T) = \sup_{v \geq 1} ess\sup E[(T - v + 1)^+ | \mathbf{Y}_1, \ldots, \mathbf{Y}_{v-1}] \tag{2.2.4}$$

over the class $\mathscr{F}_\gamma$ of all rules T satisfying the constraint $E_0(T) \geq \gamma$ on the expected duration to false alarm. He showed that as $\gamma \to \infty$, Page's (1954) CUSUM rule

$$\tau = \inf\{n : \max_{0 \leq k < n} \sum_{i=k+1}^{n} \log(f_1(\mathbf{Y}_i)/f_0(\mathbf{Y}_i)) \geq c\} \tag{2.2.5}$$

with c so chosen $E_0(\tau) = \gamma$, is asymptotically minimax in the sense that

$$\overline{E}_1(\tau) \sim \inf_{T \in \mathscr{F}_\gamma} \overline{E}_1(T) \sim (\log \gamma)/I(f_1, f_0) \tag{2.2.6}$$

where $I(f_1, f_0) = E_1\{\log(f_1(\mathbf{Y}_t)/f_0(\mathbf{Y}_t))\}$ is the Kullback-Leiber information number.

According to Lai,Liu and Xing (2009), without specifying the post-change density function in advance and supposed a multivariate exponential family

$$f_\theta(\mathbf{y}) = exp\{\boldsymbol{\theta}' y - \psi(\boldsymbol{\theta})\} \tag{2.2.7}$$

with respect to the baseline distribution so that $f_1 = f_0$, $\theta \neq 0$. Replacing the likelihood ratio statistic in the CUSUM rule by the generalized likelihood ratio (GLR)

statistic leads to the GLR rule for testing the null hypothesis of no change-point, versus the alternative hypothesis of a single change-point prior to n but not before n_0 is

$$\hat{\tau} = \inf\{n : \max_{0 \le k < n} (n-k) L(\overline{\boldsymbol{Y}}_{k+1,n}) \ge c\} \tag{2.2.8}$$

where $(n-k)L(\overline{\boldsymbol{Y}}_{k+1,n}) = \sup_\theta \sum_{i=k+1}^{n} \log(f_\theta(\boldsymbol{Y}_i)/f_0(\boldsymbol{Y}_i))$, and

$$\overline{\boldsymbol{Y}}_{m,n} = \sum_{i=m}^{n} \mathbf{Y}_i/(n-m+1), L(\boldsymbol{\mu}) = \sup_\theta \{\boldsymbol{\theta}'\boldsymbol{\mu} - \psi(\boldsymbol{\theta})\} = \{\boldsymbol{\theta}'_{\boldsymbol{\mu}}\boldsymbol{\mu} - \psi(\boldsymbol{\theta}_{\boldsymbol{\mu}}) \tag{2.2.9}$$

and $\boldsymbol{\theta}_{\boldsymbol{\mu}} = (\boldsymbol{\nabla}\psi)^{-1}(\boldsymbol{\mu})$, $\boldsymbol{\nabla}$ denotes the gradient vector of partial derivatives. Barnard (1959) gave the special case of GLR when f_θ is $N(\theta, 1)$. Siegmund and Venkatraman (1995) showed when replacing τ by $\hat{\tau}$, it is asymptotically optimal in (2.2.6) with c chosen from $E_0(\hat{\tau}) = \gamma$ and $E_0(\hat{\tau}) \sim K^{-1}c^{-1}e^c$ for a constant K as $c \to \infty$, in fact $K\sqrt{c}e^{-1}\hat{\tau}$ has a limiting exponential distribution with mean 1.

When considering the case of unknown both of the pre- and post-change density functions in advance, the GLR statistics for testing the null hypothesis of no change-point based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$, versus the alternative hypothesis of a single change-point prior to n but not before n_0 is

$$\begin{aligned} & \max_{n_0 \le k \le n} \left\{ \sup_\theta \sum_{i=1}^{k} \log f_\theta(\boldsymbol{Y}_i) + \sup_{\tilde{\theta}} \sum_{i=k+1}^{n} \log f_{\tilde{\theta}}(\boldsymbol{Y_i}) - \sup_\lambda \sum_{i=1}^{n} \log f_\lambda(\boldsymbol{Y}_i) \right\} \\ = & \max_{n_0 \le k \le n} \left\{ kL(\overline{\boldsymbol{Y}}_{1,k}) + (n-k)L(\overline{\boldsymbol{Y}}_{k+1,n}) - nL(\overline{\boldsymbol{Y}}_{1,n}) \right\} \end{aligned} \tag{2.2.10}$$

where $\sup_\lambda$ is the maximizing likelihood under the null hypothesis, and $\sup_\theta$ and $\sup_{\tilde{\theta}}$ are obtained by maximizing the likelihood under the hypothesis of a single change-point occurring at $k+1$. Let

$$g(\alpha, x, y) = \alpha L(x) + (1-\alpha)L(y) - L(\alpha x + (1-\alpha)y). \tag{2.2.11}$$

the GLR rule for detecting a change in θ when the pre- and post-parameters are unknown is

$$\widehat{N} = \inf\left\{ n > n_0 : \max_{n_0 \le k \le n} \left\{ \sup_\theta \sum_{i=1}^{k} \log f_\theta(\boldsymbol{Y}_i) + \sup_{\tilde{\theta}} \sum_{i=k+1}^{n} \log f_{\tilde{\theta}}(\boldsymbol{Y_i}) - \sup_\lambda \sum_{i=1}^{n} \log f_\lambda(\boldsymbol{Y}_i) \right\} \ge c \right\} \tag{2.2.12}$$

2.3 Detection Method Based on An Extension of Shiryaev's Bayesian Single Change-Point Model

Based on the model of Lai and Xing (2010), the generalization form of the extension of Shiryaev's Bayesian change-point model and detection rule is proposed as following: for the multiparameter exponential family (2.2.7), let π be a prior density function on $\boldsymbol{\Theta} := \{\boldsymbol{\theta} : \int e^{\boldsymbol{\theta' Y}} d\omega(\boldsymbol{Y}) < \infty\}$ given by

$$\pi(\boldsymbol{\theta}; a_0, \boldsymbol{\mu_0}) = c(a_0, \boldsymbol{\mu_0}) \exp\{a_0 \boldsymbol{\mu_0'\theta} - a_0 \psi(\boldsymbol{\theta})\}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{2.3.1}$$

where $1/c(a_0, \boldsymbol{\mu_0}) = \int_{\boldsymbol{\Theta}} \exp\{a_0 \boldsymbol{\mu_0'} \boldsymbol{\theta} - a_0 \psi(\boldsymbol{\theta})\} d\theta$, and $\boldsymbol{\mu_0} \in (\boldsymbol{\nabla}\psi)(\boldsymbol{\Theta})$. Based on Diaconis and Ylvisaker (1979, p.274), the posterior density of $\boldsymbol{\theta}$ given the observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ drawn from $f_{\boldsymbol{\theta}}$ is

$$\pi(\boldsymbol{\theta}; a_0 + m, (a_0 \boldsymbol{\mu_0} + \sum_{i=1}^{m} \boldsymbol{Y}_i)/(a_0 + m)) \tag{2.3.2}$$

Moreover,

$$\int_{\boldsymbol{\Theta}} f_{\boldsymbol{\theta}}(\boldsymbol{Y}) \pi(\boldsymbol{\theta}; a, \boldsymbol{\mu}) d\theta = \frac{c(a, \boldsymbol{\mu})}{c(a+1, (a\boldsymbol{\mu} + \boldsymbol{Y})/(a+1))} \tag{2.3.3}$$

Suppose the change-time v and the pre- and post-change values of parameter are unknown, and for $t < v$, the parameter $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ for $t \geq v$. Following Shiryaev (1963,1978), Lai and Xing (2010) use the Bayesian approach that assumes v to be geometric with parameter p but constrained to be lager than n_0, and $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ are i.i.d of (2.3.1), which is also independent of v. The modification of Shiryaev's rule is as follows. Let $\mathscr{F}_t$ denote the σ-field generated by $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t$. Let $\pi_{0,0} = c(a_0, \boldsymbol{\mu}_0)$ and $\pi_{i,j} = c(a_0 + j - i + 1, a_0 \boldsymbol{\mu}_0 + \sum_{t=i}^{j} \boldsymbol{Y}_t)/(a_0 + j - i + 1))$. For $n_0 < i < n$,

$$P\{v = i | \mathscr{F}_n\} \propto p(1-p)^{i-1} \pi_{0,0}^2 / \pi_{1,i-1} \pi_{i,n}, \quad P\{v > n | \mathscr{F}_n\} \propto p(1-p)^n \pi_{0,0} / \pi_{1,n} \tag{2.3.4}$$

and we express $P(n_0 < v \leq n | \mathscr{F}_n) = \sum_{i=n_0+1}^{n} P\{v = i | \mathscr{F}_n\}$ in terms of $\pi_{i,j}$. Therefore, the format of Shiryaev's stopping rule for unknown pre-and post-change param-

eters can be given with replacement of $R_{p,n} = \sum_{i=n_0+1}^{n} \pi_{0,0}\pi_{1,n}/\{(1-p)^{n-i}\pi_{1,i}\pi_{i,n}\}$ in (2.2.2). And it can be modified to

$$N = \inf\{n > n_p : P(v \leq n | v \geq n - k_p, \mathscr{F}_n) \geq \eta_p\} \tag{2.3.5}$$

Since

$$P(v \leq n | v \geq n^* - k_p, \mathscr{F}_n) = \sum_{i=n-k_p}^{n} P(v = i|\mathscr{F}_n) / \left\{ \sum_{i=n-k_p}^{n} P(v = i|\mathscr{F}_n) + P(v > n|\mathscr{F}_n) \right\} \tag{2.3.6}$$

we can get

$$N = \inf \left\{ n > n_p : \sum_{i=n-k_p}^{n} \frac{\pi_{0,0}\pi_{1,n}}{(1-p)^{n-i+1}\pi_{1,i-1}\pi_{i,n}} \geq \gamma_p \right\} \tag{2.3.7}$$

Specifically, based on the model in Xing et al.(2012), assuming the structural breaks in credit rating generator matrices follow a Poisson process $\{N_\Lambda(t); t \geq 0\}$ with constant rate η, which means the duration between two adjacent structural breaks follows an exponential distribution with mean $1/\eta$. The generator matrices between two adjacent structural breaks are constant and at time t, the generator matrix is $\Lambda(t) = Q_{N_\Lambda(t)}$, in which $Q_1, Q_2, \ldots$ are independent and identically distributed random generator matrices with the off-diagonal elements $\lambda^{(i,j)}$ following

independently Gamma(α_{ij}, β_i) prior distribution

$$g(\lambda^{(i,j)}) = \frac{\beta_i^{\alpha_{ij}}}{\Gamma(\alpha_{ij})}[\lambda^{(i,j)}]^{\alpha_{ij}-1}\exp(-\lambda^{(i,j)}\beta_i), \quad (i,j) \in \mathcal{K}, \tag{2.3.8}$$

where $\mathcal{K} = \{(i,j)|i \neq j, 1 \leq i \leq K-1, 1 \leq j \leq K\}$. The elements in the last row are always zero, which presents the rating migrations from default to other class. Also, the assumptions suggest that $\Lambda(t)$ follows a compound Poisson process with rate η, and for the time period (s,t), the time-dependent credit rating transition matrix $P(s,t)$ can be characterized in the following way.

If there are M structural breaks in the period (s,t) and assuming the observed M time changes satisfy $s < \tau_1 < \cdots < \tau_M < t$, then the transition matrix in the period (s,t) are

$$P(s,t) = \prod_{k=1}^{M+1} P(\tau_{k-1}, \tau_k) = \prod_{k=1}^{M+1} \exp\left(\int_{\tau_{k-1}}^{\tau_k} \Lambda(u)du\right) = \prod_{k=1}^{M+1} \exp\left[(\tau_k - \tau_{k-1})\Lambda(\tau_k-)\right]$$

where $\tau_0 = s, \tau_{M+1} = t$.

If there is no changes in the generator $\Lambda(t)$ over the period (s,t),the transition matrix $P(s,t)$ becomes homogeneous

$$P(s,t) = \exp\left(\int_s^t \Lambda(u)du\right) = \exp\left[(t-s)\Lambda(t-)\right] \tag{2.3.9}$$

In this case, over the period, due to the assumption of the elements $\Lambda(t)$ following

the conjugate Gamma priors (2.3.8), we can calculate the posterior distribution of $\Lambda(t)$ given transition history as follows. Suppose there are n realizations of a Markov chain with generator matrices $\Lambda(t)$. Here, in the period (s,t), let $\mathcal{Y}_{s,t}$ be the observed transitions, $K_{s,t}^{(i,j)}$ be the number of transitions from class i to class j, $S_{s,t}^{(i)}$ be the total time spending in class i, and $\lambda_{s,t}^{(i,j)}$ be the ijth entry in the generator $\Lambda(t)$. Then given the constant $\Lambda(t)$, the likelihood of $\mathcal{Y}_{s,t}$ can be expressed as

$$\exp\left\{\sum_{i=1}^{K}\left[\sum_{j\neq i}K_{s,t}^{(i,j)}\log\lambda_{s,t}^{(i,j)}-(\sum_{j\neq i}\lambda_{s,t}^{(i,j)}+1-K)S_{s,t}^{(i)}\right]\right\}\propto\prod_{i\neq j}(\lambda_{s,t}^{(i,j)})^{K_{s,t}^{(i,j)}}e^{-\lambda_{s,t}^{(i,j)}S_{s,t}^{(i)}} \tag{2.3.10}$$

see Küchler and Sørensen (1997, p. 26). Combining this with the prior in (2.3.8), the posterior distribution of $\lambda_{s,t}^{(i,j)}$ given $\mathcal{Y}_{s,t}$ is Gamma $(K_{s,t}^{(i,j)}+\alpha_{ij}, S_{s,t}^{(i)}+\beta_i)$. Hence we can get the estimation of the element $\lambda_{s,t}^{(i,j)}$ by the posterior mean of the Gamma distribution, i.e., $\widehat{\lambda}_{s,t}^{(i,j)}=(K_{s,t}^{(i,j)}+\alpha_{ij})/(S_{s,t}^{(i)}+\beta_i)$.

Also, the proposed model is based on the assumption that firm's rating transition from state i to state$j$ over period (s,t) are conditional independent given the generator matrix in the period of (s,t).

In Xing et al.(2012), the posterior distribution of $\Lambda(t_l)=(\lambda_{t_{l-1},t_l}^{(i,j)})_{(i,j)}\in\mathcal{K}$ given $\mathcal{Y}_{(0,T)}$ can be characterized as following: let $R_l=\max\{t_{m-1}|I_m=1, m\leq l\}$, i.e., R_l represents the most recent structural break up to time t_{l-1}. From the above assumption, given $R_l=t_{m-1}$, we can get the conditional distribution of $\lambda_{t_{m-1},t_l}^{(i,j)}$,

which is Gamma$(K^{(i,j)}_{t_{m-1},t_l}+\alpha_{ij}, S^{(i)}_{t_{m-1},t_l}+\beta_i)$. Let $p_{m,l} = P(R_l = t_{m-1}|\mathcal{Y}_{t_{m-1},t_l})$. Then the posterior distribution of $\lambda^{(i,j)}_{t_{l-1},t_l}$ given $\mathcal{Y}_{(0,t_l)}$ can be expressed as a mixture of Gamma distributions,

$$\lambda^{(i,j)}_{t_{l-1},t_l}|\mathcal{Y}_{(0,t_l)} \sim \sum_{m=1}^{l} p_{m,l} Gamma(K^{(i,j)}_{t_{m-1},t_l}+\alpha_{ij}, S^{(i)}_{t_{m-1},t_l}+\beta_i) \tag{2.3.11}$$

where the mixture weight is recursively calculated by $p_{m,l} = p^*_{m,l}/\sum_{m=1}^{l} p^*_{m,l}$, in which

$$p^*_{m,l} = \begin{cases} p f_{l,l}/f_{0,0} & m = l, \\ (1-p) p_{m,l-1} f_{m,l}/f_{m,l-1} & m < l \end{cases} \tag{2.3.12}$$

The terms $f_{m,l}$ and $f_{0,0}$ in (2.3.12) are expressed as follows:

$$f_{m,l} = \prod_{i,j\in\mathcal{K}} \Gamma(K^{(i,j)}_{t_{m-1},t_l}+\alpha_{ij})/(S^{(i)}_{t_{m-1},t_l}+\beta_i)^{(K^{(ij)}_{t_{m-1},t_l}+\alpha_{ij})}, \tag{2.3.13}$$

$$f_{0,0} = \prod_{i,j\in\mathcal{K}} \Gamma(\alpha_{i,j})/\beta_i^{\alpha_{i,j}} \tag{2.3.14}$$

The proof of the above posterior distributions given in the appendix A in Xing et al.(2012) can be briefly expressed as follows:

Let $f(.|\mathcal{Y}_{(0,t_l]})$ denote the density function of $\lambda^{(i,j)}_{t_{l-1},t_l}$ given $\mathcal{Y}_{(0,t_l]}$. We have

$$\begin{aligned} f(\lambda^{(i,j)}_{t_{l-1},t_l}|\mathcal{Y}_{(0,t_l]}) &\propto f(\lambda^{(i,j)}_{t_{l-1},t_l}, \mathcal{Y}_{(t_{l-1},t_l]}|\mathcal{Y}_{(0,t_{l-1}]}) \\ &= pf(\lambda^{(i,j)}_{t_{l-1},t_l}, \mathcal{Y}_{(t_{l-1},t_l]}|\mathcal{Y}_{(0,t_{l-1}]}, I_{l=1}) + (1-p)f(\lambda^{(i,j)}_{t_{l-1},t_l}, \mathcal{Y}_{(t_{l-1},t_l]}|\mathcal{Y}_{(0,t_{l-1}]}, I_{l=0}) \end{aligned}$$

in which

$$\begin{aligned} pf(\lambda^{(i,j)}_{t_{l-1},t_l}, \mathcal{Y}_{(t_{l-1},t_l]}|\mathcal{Y}_{(0,t_{l-1}]}, I_{l=1}) &= p^*_{l,l} f(\lambda^{(i,j)}_{t_{l-1},t_l}|\mathcal{Y}_{(0,t_l]}, I_{l=1}) \\ &= p^*_{l,l} Gamma(K^{(i,j)}_{t_{l-1},t_l} + \alpha_{ij}, S^{(i)}_{t_{l-1},t_l} + \beta_i) \end{aligned}$$

where

$$\begin{aligned} p^*_{l,l} &= pf(\mathcal{Y}_{(t_{l-1},t_l]}|\mathcal{Y}_{(0,t_{l-1}]}, I_{l=1}) \\ &= p\int f(\mathcal{Y}_{(t_{l-1},t_l]}|\lambda^{(i,j)}_{t_{l-1},t_l}) g(\lambda^{(i,j)}_{t_{l-1},t_l} d\lambda^{(i,j)}_{t_{l-1},t_l} \\ &= pf_{l,l}/f_{0,0} \end{aligned}$$

and

$$\begin{aligned} & (1-p)f(\lambda^{(i,j)}_{t_{l-1},t_l}, \mathcal{Y}_{(t_{l-1},t_l]}|\mathcal{Y}_{(0,t_{l-1}]}, I_{l=0}) \\ = \; & (1-p)\sum_{m=1}^{l-1} P(R_{l-1} = t_{m-1}|\mathcal{Y}_{(0,t_{l-1}]}, I_{l=0}) \times \\ & f(\lambda^{(i,j)}_{t_{l-1},t_l}, \mathcal{Y}_{(t_{l-1},t_l]}|R_{l-1} = t_{m-1}, \mathcal{Y}_{(0,t_{l-1}]}, I_l = 0) \\ = \; & \sum_{m=1}^{l-1} p^*_{m,l} f(\lambda^{(i,j)}_{t_{l-1},t_l}|R_{l-1} = t_{m-1}, \mathcal{Y}_{(0,t_l]}, I_{l=0}) \end{aligned}$$

$$= \sum_{m=1}^{l-1} p^*_{m,l} Gamma(K^{(i,j)}_{t_{m-1},t_l} + \alpha_{ij}, S^{(i)}_{t_{m-1},t_l} + \beta_i)$$

where

$$\begin{aligned} p^*_{m,l} &= (1-p)p_{m,l-1} f(\mathcal{Y}_{(t_{l-1},t_l]} | R_{l-1} = t_{m-1}, \mathcal{Y}_{(0,t_{l-1}]}, I_l = 0) \\ &= (1-p)p_{m,l-1} \frac{f(\mathcal{Y}_{t_{m-1},t_l}, R_l = t_{m-1})}{f(\mathcal{Y}_{t_{m-1},t_{l-1}}, R_{l-1} = t_{m-1})} \\ &= (1-p)p_{m,l-1} f_{m,l}/f_{m,l-1} \end{aligned}$$

Combining the above model with the rule in (2.3.5), let $\mathscr{F}_t$ denote the σ-field generated by $I_1, I_2, \ldots, I_t$. Let m be the time of change point, the associated detection rule for testing the null hypothesis of no change-point, versus the althernative hypothesis of a single change-point to n but not before n_0 is

$$N = \inf \left\{ n > n_p : P(m \leq n | m \geq n - k_p, \mathscr{F}_n) \geq \eta_p \right\} \tag{2.3.15}$$

and

$$P\{m = r|\mathscr{F}_n\} \propto p(1-p)^{r-1} f_{1,r-1} f_{r,l}/f^2_{0,0}, P\{m > n|\mathscr{F}_n\} \propto p(1-p)^n f_{1,l}/f_{0,0} \tag{2.3.16}$$

we can use (2.3.16) to rewrite (2.3.15) in the form

$$N = \inf\left\{n > n_p : \sum_{r=n-k_p}^{n} \frac{f_{1,r-1} f_{r,n}}{(1-p)^{n-r+1} f_{0,0} f_{1,n}} \geq \gamma_p\right\} \tag{2.3.17}$$

2.4 A Sequential Detection Rule For Multiple Structural Breaks

The detection rule proposed by Lai,Liu and Xing (2009) focus on the Bayesian change-point model which does not have to retain all past observations and the posterior probability can be calculated by explicit recursive formulas. We can summarize below.

Consider the same multiparameter model in section 2.2, which expressed in (2.2.7),(2.3.1),(2.3.2) and (2.3.3). Suppose the parameter vector $\boldsymbol{\theta_t}$ may undergo occasional changes for $t > 1$ and as in Chernoff and Zacks (1964), the indicator variables

$$I_t := \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \tag{2.4.1}$$

are independent Bernoulli random variables with $P(I_t = 1) = p$. Generalizing Yao (1984), when a parameter changing occurred at time t (i.e., $I_t = 1$), the changed parameter $\boldsymbol{\theta}_t$ is assumed to be sampled from π. Denote K_t as the most recent change-time up to t, i.e., $K_t = \max\{s \leq t : I_s = 1\}$, and $f(|)$ as the conditional

densities

$$f(\boldsymbol{\theta}_t|\mathcal{Y}_t) = \sum_{i=1}^{t} p_{it} f(\boldsymbol{\theta}_t|\mathcal{Y}_{i,t}, K_t = i) \tag{2.4.2}$$

where $p_{it} = P(K_t = i|\mathcal{Y}_t)$. Following (2.3.1), we have

$$f(\boldsymbol{\theta}_t|\mathcal{Y}_{i,t}, K_t = i) = \pi(\boldsymbol{\theta}_t; a_0 + t - i + 1, \overline{\boldsymbol{Y}}_{i,t}) \tag{2.4.3}$$

where $\overline{\boldsymbol{Y}}_{i,j} = (a_0\boldsymbol{\mu}_0 + \sum_{k=i}^{j} \boldsymbol{y}_k)/(a_0 + j - i + 1)$ for $j \geq i$. Combining (2.4.2) with (2.4.3), we get

$$f(\boldsymbol{\theta}_t|\mathcal{Y}_t) = \sum_{i=1}^{t} p_{it} \pi(\boldsymbol{\theta}_t; a_0 + t - i + 1, \overline{\boldsymbol{Y}}_{i,t}) \tag{2.4.4}$$

Note, $\sum_{i=1}^{t} p_{it} = 1$, the recursive formula can be characterized

$$p_{it} \propto p_{it}^* = \begin{cases} pf(\boldsymbol{y}_t|I_t = 1) & i = t, \\ (1-p)p_{i,t-1} f(\boldsymbol{y}_t|\mathcal{Y}_{i,t-1}, K_t = i) & i \leq t-1 \end{cases} \tag{2.4.5}$$

When combining $f(\boldsymbol{y}_t|\mathcal{Y}_{i,t-1}, K_t = i) = \int f_{\boldsymbol{\theta}_t}(\boldsymbol{y}) f(\boldsymbol{\theta}_t|\mathcal{Y}_{i,t-1}, K_t = i) d\boldsymbol{\theta}_t$ with (2.2.7),(2.3.3) and (2.4.2) yields

$$p_{it}^* = \begin{cases} p\frac{\pi_{0,0}}{\pi_{t,t}} & i = t, \\ (1-p)p_{i,t-1}\frac{\pi_{i,t-1}}{\pi_{i,t}} & i < t \end{cases} \tag{2.4.6}$$

where $\pi_{0,0} = c(a_0, \boldsymbol{\mu}_0)$ and $\pi_{i,j} = c(a_0 + j - i + 1, \overline{\boldsymbol{Y}}_{i,j})$.Note that $\sum_{t-m}^{t} p_{it}$ is the

posterior mean number of change-points in the time interval between $t-m$ and t given $\mathcal{Y}_t$.

Following the assumption proposed in Shiryaev (1978), assuming a loss of 1 or $c(N-v)$ if $N < v$ or $N \geq v$ when the stopping time is N. Shiryaev has used optimal stopping theory to show that there exists a constant $\delta_{p,c}$ such that the threshold rule stopping at time n, when

$$P(v \leq n|\mathcal{Y}_n) \geq \delta_{p,c} \tag{2.4.7}$$

is optimal.

Note that $\{k \leq v \leq n\} = \cup_{i=k}^{n}\{K_n = i\}$. Based on the more general model which allows unknown pre- and post-change parameters and multiple change-points, Lai,Liu and Xing (2009) has proposed a modified Shiryaev's rule (2.4.7) that a change-point occurred in the time interval between $n_k(p)$ and n, declaring at time n if

$$\sum_{i=n-k(p)}^{n} p_{in} \geq \eta_p \tag{2.4.8}$$

Consider the model proposed by Xing et al.(2012) and apply the above detection

rule. In the time interval between $l - k(p)$ and l, if

$$L = \inf \left\{ l > l_p : \sum_{r=l-k(p)}^{l} p_{r,l} \geq \eta_p \right\} \tag{2.4.9}$$

for suitably chosen η_p, and $p_{r,l} = p^*_{r,l} / \sum_{r=1}^{l} p^*_{r,l}$ and $p^*_{r,l}$ is defined in 2.3.12.

Chapter 3

Numerical Studies and Results

3.1 Simulation Studies-Change Points Detection

This section presents the simulation studies of the four detection methods proposed in section 2. For simplicity, we only consider a 2-state Markov Chain as the transition matrix for the credit rating dynamics, and let the second state as the default state. A more complicated model with 4-state Markov Chain will be discussed in next section. All the change point process is generated following the credit rating model in Xing et al.(2012). Four different scenarios are considered in this study.

3.1.1 Data Generation

For the 2-state Markov chain model, we assume that there are 2 rating categories 1 and 2 observed over the sample time period $(0, 3)$, where state 2 is absorbing state and can not transfer to state 1. We consider an evenly spaced partition in the time period $(0, 3)$, $0 = t_0 < t_1 < \cdots < t_L = 3$ and let each time period be 0.01, so there are totally 300 time periods in our model. In each scenario, 1000 samples (simulations) of size n=1000 firms (suppose there are 1000 firms) are generated to detect the change-point time.

The parameter of structural break in a 2-state Markov chain model is generated as

$$\Lambda(s) = \begin{pmatrix} -\lambda(s) & \lambda(s) \\ 0 & 0 \end{pmatrix}$$

and the steps are as following:

Step 1. Pre-specify the change-point time. In our study, do avoid the impact from the previous change-point in the sequential change-point model, we suppose that there are more than 50 time periods between adjacent structural breaks. In our data, the structural breaks are no more than 3 for each sample.

Step 2. From (2.3.9), let $\Delta s = 0.01$, and combining Λ, we can get the rating

transition matrix is as following:

$$\begin{aligned} P(s, s+\Delta s) &= \exp\left(\int_s^{s+\Delta s} \Lambda(u)du\right) \\ &\approx \exp[\Lambda(s)\Delta s] \\ &= \sum_{k=0}^{\infty} \frac{(\Lambda(s)\Delta s)^k}{k!} \\ &= \begin{pmatrix} \sum_{k=0}^{\infty} \frac{(-\lambda(s)\Delta s)^k}{k!} & 1 - \sum_{k=0}^{\infty} \frac{(-\lambda(s)\Delta s)^k}{k!} \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \exp(-\lambda(s)\Delta s) & 1 - \exp(-\lambda(s)\Delta s) \\ 0 & 1 \end{pmatrix} \end{aligned}$$

In this study, we let the absorbing firms less than 100 until $t_L = 3$, in other words, there are less than 100 firms transferred to state 2 in the total time period $(0, 3)$.

3.1.2 Scenarios Statement

The four scenarios are: no change-point model, single change-point model, two change points model, and multiple change points model with change-point process following the model in Xing et al.(2012).

Scenario 1. No change-point model. The data are generated from a constant parameter model with parameter $\lambda = 0.01$ based on the rating transition matrix given in section 3.1.1.

Scenario 2. Single change-point model. The data are generated from a single change-point model with change-point at $t = 2$, and $\lambda_0 = 0.015$ for $0 \leq t < 2$, $\lambda_1 = 10^{-5}$ for $2 \leq t \leq 3$.

Scenario 3. Two change points model. The data are generated from a two change points model with the first change-point at $t = 1$, the second change-point at $t = 2$ and $\lambda_0 = 0.01$ for $0 \leq t < 1$, $\lambda_1 = 0.012$ for $1 \leq t < 2$, $\lambda_2 = 0.015$ for $2 \leq t \leq 3$.

Scenario 4. Multiple change points model. The data are generated based on the structural break model in Xing et al.(2012). For step 1, we have to simulate a Poisson process with a constant rate η during the time period $(0, 3)$, such that $\{\tau_m - \tau_{m-1}\} \sim \exp(\eta)$, where τ_m is the time of change-point occurred. Here, we let $\eta = 0.00105$. For step 2, According to $\Lambda(t) = Q_{N_\Lambda(t)}$ and the assumption of the prior distribution followed a Gamma(α_{ij}, β_i), the generator matrices of a 2-state Markov chain model can be characterized as

$$\Lambda(\tau_m) = \begin{pmatrix} -\lambda(\tau_m) & \lambda(\tau_m) \\ 0 & 0 \end{pmatrix}$$

where $\lambda(\tau_m) \sim$ Gamma(α_{12}, β_1). Here, we let $\alpha_{12} = 0.9067, \beta_1 = 100$. From the formula of transition probability, the transition matrix with a change-point occurring

during one time period (t_l, t_{l+1}) is:

$$P(t_l, t_{l+1}) = \begin{pmatrix} e^{-\lambda(\tau_{i-1})(\tau_i - t_l)} & 1 - e^{-\lambda(\tau_{i-1})(\tau_i - t_l)} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-\lambda(\tau_i)(t_{l+1} - \tau_i)} & 1 - e^{-\lambda(\tau_i)(t_{l+1} - \tau_i)} \\ 0 & 1 \end{pmatrix}$$

3.1.3 Simulation Results

In our study, given α=5%, 3%, 2%, we choose the threshold θ in(2.1.3), c in(2.2.12), γ_p in(2.3.17) and η_p in (2.4.9) by using Monte Carlo simulation under the null hypothesis of no change-point based on the data of Scenario 1 with constant λ of the transition matrix for each of the four detection rules.

For Shewhart Control Charts detection method, we choose $\alpha = 0.2$ in (2.1.3) and figure 3.1 shows one sample of the $\mathbf{Z}_t$'s (2.1.3) value as time changes on a single change-point simulation based on a 2-state Markov chain model. From the path of $\mathbf{Z}_t$'s value, we can clearly see the change point time around $t = 200$.

Table 3.1 compares the performance of the 4 methods for the single change-point model with change-point time at t=200. We choose $n_p = 5$ and $k_p = 30$ in the detection rule of the extension of Shiryaev's Bayesian change-point model (2.3.17) and the sequential detection rule for multiple structural breaks (2.4.9). Once a change-point time is detected, the data before this time is not considered for the next time point detection. For each detection rule, we list the expected detection

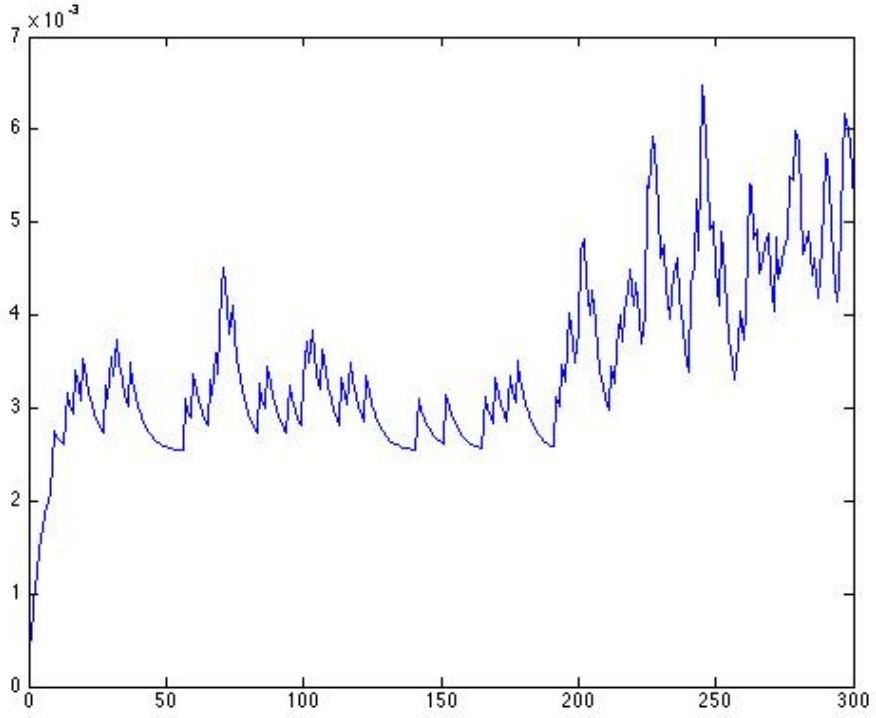

Figure 3.1: EWMA Control Charts detection method: value of $\mathbf{Z}_t$ as time changes on a 2-state Markov chain model

delay $E(N-m)^+$ based on 1000 simulations, and standard errors are given in the parentheses. To simplify the written of method's name, we label rule 1 for the Shewhart control chart detection method, label rule 2 for the generalized likelihood ratio (GLR) detection rule on single change-point model, label rule 3 for the detection based on an extension of Shiryaev's Baye's single change-point model, and label rule 4 for the sequential detection method for multiple structure breaks.

Table 3.1: Expected detection delays for Scenario 2 (single change-point model)

α	Detection rule	C-P time t = 200	Detected number (based on 1000)
0.05	Rule-1(2.1.3)	9.55(43.43)	811
	Rule-2(2.2.12)	30.78(20.20)	986
	Rule-3(2.3.17)	7.23(7.69)	995
	Rule-4(2.4.9)	9.63(6.88)	979
0.03	Rule-1	8.89(52.44)	931
	Rule-2	33.85(21.67)	977
	Rule-3	14.21(11.24)	982
	Rule-4	13.20(9.16)	955
0.02	Rule-1	10.31(72.57)	969
	Rule-2	36.66(22.19)	961
	Rule-3	14.28(11.44)	978
	Rule-4	13.23(9.30)	950

In table 3.1, the second column is the expected detection delay of $t = 200$, and the according standard error is in the parentheses. For 1000 simulations, the total number of change-points is 1000 and the number of change-point that we can detect is listed in the third column. From this comparison, we can see although rule 1 shows a good expected detection delay, but has very large variance. The expected detection delay in rule 2 is large. Only rule 3 and rule 4 have both acceptable expected detection delay and low standard errors.

For the detection rule 2, rule 3 and rule 4, we have a more detailed analysis based on scenario 3-two change points model and scenario 4-multiple change points based on the structural break model in Xing et al.(2012). Table 3.2 and table 3.3 give the detection results for this study.

Table 3.2: Expected detection delays for Scenario 3 (two change points model)

α	Rule	C-P time t_1= 100	C-P time t_2=200	Detected number (based on 2000)
0.05	Rule-2	17.95(12.68)	38.42(19.51)	1994
	Rule-3	7.31(6.38)	16.62(4.54)	1980
	Rule-4	7.34(6.84)	16.58(4.43)	1992
0.03	Rule-2	20.22(13.50)	44.48(16.27)	1984
	Rule-3	7.38(5.53)	17.99(4.66)	1976
	Rule-4	7.54(7.11)	18.11(4.56)	1984
0.02	Rule-2	22.56(15.01)	47.62(18.85)	1970
	Rule-3	7.57(5.84)	19.13(4.72)	1960
	Rule-4	7.89(8.85)	19.05(4.63)	1971

Table 3.3: Expected detection delays for Scenario 4 (multiple change points model based on Xing et al.(2012))

α	Rule	C-P time $t_1 = \tau_1$	C-P time $t_2 = \tau_2$	C-P time $t_3 = \tau_3$	Detected number (based on 1731)
0.05	Rule-2	12.85(35.91)	16.58(28.98)	62.09(33.57)	1030
	Rule-3	8.98(26.21)	25.31(32.27)	27.79(26.34)	1009
	Rule-4	8.25(24.37)	24.59(34.70)	28.77(27.14)	984
0.03	Rule-2	12.69(36.45)	16.25(30.07)	29.87(∞)	1002
	Rule-3	8.79(26.64)	24.39(31.03)	28.68(27.20)	976
	Rule-4	8.68(26.43)	23.76(34.40)	28.77(27.14)	958
0.02	Rule-2	13.48(37.87)	14.39(27.68)	47.87(∞)	986
	Rule-3	8.95(27.56)	25.09(31.85)	29.99(26.25)	957
	Rule-4	8.61(26.99)	22.76(32.44)	29.23(26.65)	941

Here, except for the expected detection delay and the standard errors, another important measurement indicator should be considered which we call detected number. First,we denote two kinds of detection error. One is the error not showing change-point when it does have, and the other one falsely shows the change-point when it does not have. In our study, usually, the second kind of error happens before the time of the first change-point happens (we can call it pre-detection in this situation). The detected number in the right list of the tables is the theoretical change points number we deducts the first kind of errors, and the theoretical number of change-points in 1000 simulations is given in the parentheses.

In table 3.2 (two change points model), the change-point time is $t_1 = 100$, $t_2 = 200$, and the total number of change-points of 1000 simulations is 2000. It is clearly to see that detection rule-2 has a worse expected detection delay and standard error comparing with the other two rules, and the expected detection delay and the standard errors of rule-3 and rule-4 are almost the same, but rule-4 has a little bit more detected numbers than rule-3. For a comprehensive analysis of this table, the result of detection rule-4 is more acceptable.

Table 3.3 shows the result for the multiple change points based on scenario 4, we can get almost the same result with table 3.2. For this multiple change points model, the correct detected number of change points is less than other scenarios.

To more simply and clearly compare for detection rule 2, rule 3 and rule 4, we

also plots some figures to present the results.

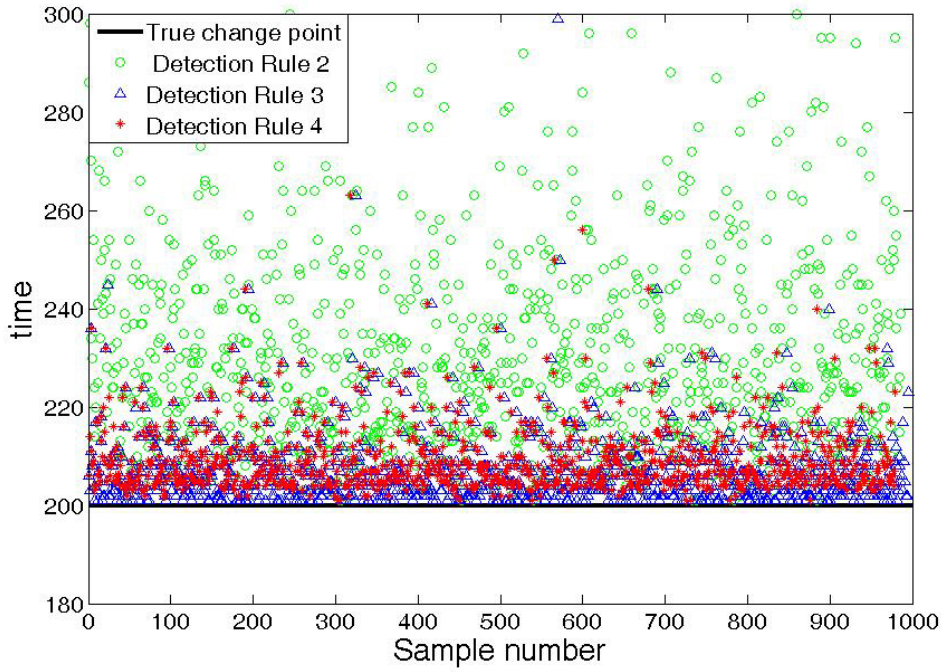

Figure 3.2: Plots of detected change points given $\alpha = 0.05$ for scenario 2: single change-point model with change-point at $t = 200$ in rule 2, rule 3 and rule 4

Figure 3.2 plots the detected time of rule 2, rule 3 and rule 4 for scenario 2. We can see that the blue dots which represent the result of rule 3 are closest to the line of change-point time at $t = 200$, and the green dots are farthest among the three rules. In the single change-point model, rule 3 and rule 4 have a much better detection result than rule 2.

Figure 3.3 is the same plot as 3.2 for scenario 3 and we can get the same result with table 3.2. In this figure, the blue dots which represent rule 3 and the red dots

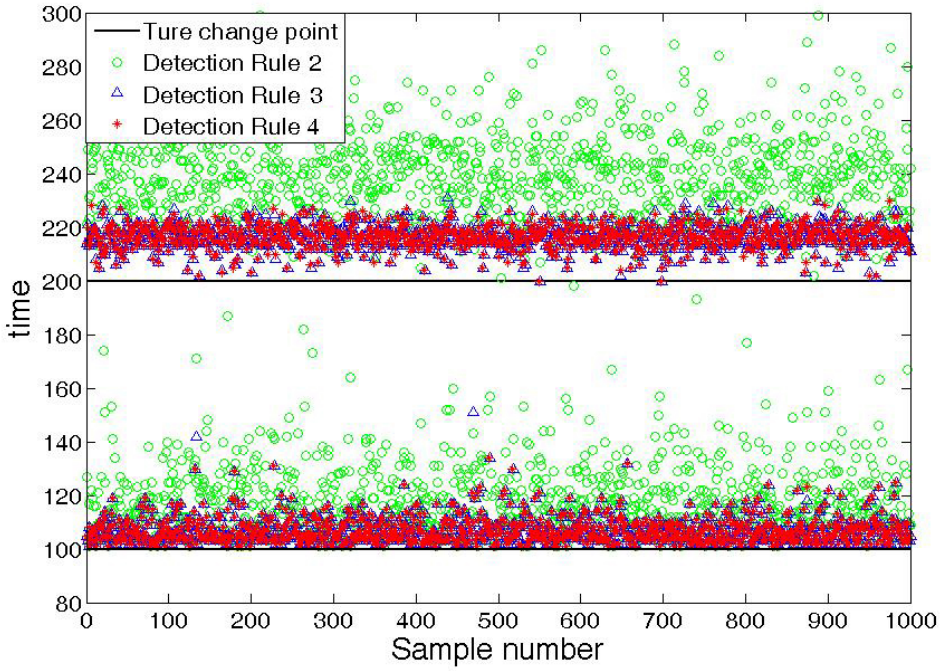

Figure 3.3: Plots of detected change points given $\alpha = 0.05$ for scenario 3: two change points at $t_1 = 100$, $t_2 = 200$

which represent rule 4 show very acceptable results for both change time at $t_1 = 100$ and $t_2 = 200$. The green dots for rule 2 is also the farthest among the three rules.

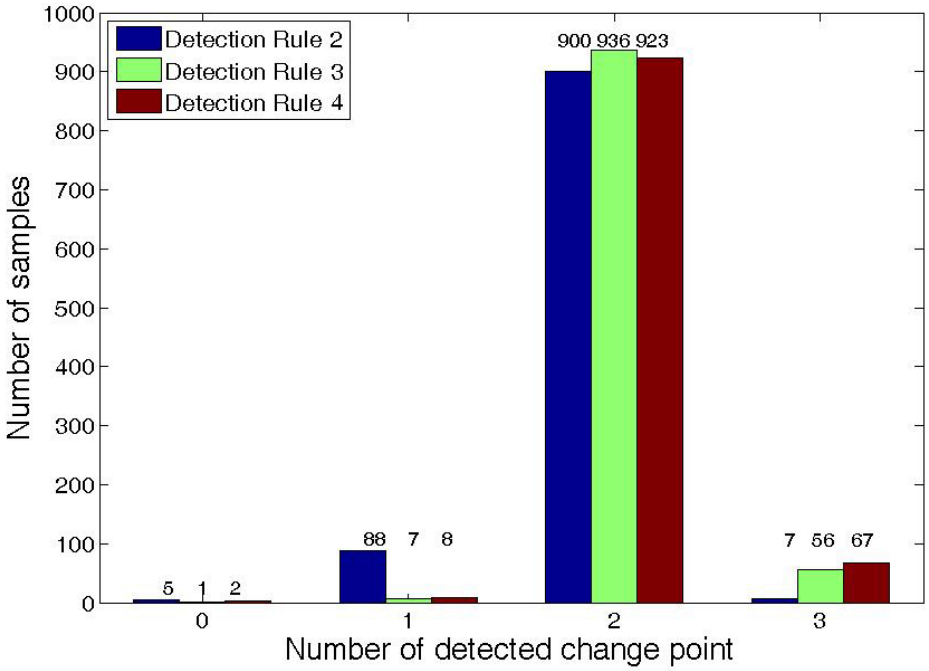

Figure 3.4: Detected number of 1000 simulations for scenario 3: two change points model. 0=detected none, 1=detected 1 C-P, 2=detected 2 C-Ps, 3=detected 3 C-Ps

Figure 3.4 is the histogram of the detected number of change-points of 1000 simulations for scenario 3. From this figure, rule 3 (green bar) has the most number of 2 detected change points, which means the correct detection. All the three detection rules (rule 2, rule 3 and rule 4) have a more than 90% correct detection.

3.2 Simulation Studies-Critical Value Determined

In last section, we have roughly detected both single change-point model and multiple change points model based a 2-state Markov chain transition matrix. In this section, we focus on the detection based on a more theoretical threshold. In this study, the data with no change-point in 1000 simulations is used to determine the critical value, which can be seen as the theoretical threshold. Here, we consider a more complicated Markov chain model with 4 transition states and a longer time periods of 500.

3.2.1 Data Generation and Critical Value Determined Procedure

Predetermine G_0 as the prior, and G_0 is:

$$(\alpha_{ij}) = \begin{pmatrix} . & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & . & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & . & \alpha_{34} \\ . & . & . & . \end{pmatrix} \qquad (\beta_i) = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & . \end{pmatrix}$$

Step 1. Generate $\Lambda_1, \Lambda_2, \ldots, \Lambda m$, $m = 1000$.

$$\Lambda = \begin{pmatrix} -\sum_{j=2}^{4} \lambda_{1j} & \lambda_{12} & \lambda_{13} & \lambda_{14} \\ \lambda_{21} & -\sum_{j=1,j\neq 2}^{4} \lambda_{2j} & \lambda_{23} & \lambda_{24} \\ \lambda_{31} & \lambda_{32} & -\sum_{j=1,j\neq 3}^{4} \lambda_{3j} & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where λ_{ij} follows Gamma(α_{ij}, β_i) prior distribution with

$$g(\lambda_{(ij)}) = \frac{\beta_i^{\alpha_{ij}}}{\Gamma(\alpha_{ij})} \times [\lambda_{(ij)}^{\alpha_{ij}-1}] \times exp(-\lambda_{(ij)}\beta_i), i, j \in K$$

Step 2. From section 3.1.1, we know that transition matrix $P \approx exp(\delta t \times \Lambda)$. Let $\delta t = \frac{1}{12}$, which is according to the monthly credit data, we have $P_1 = exp(\frac{1}{12} \times \Lambda_1), \ldots, P_m = exp(\frac{1}{12} \times \Lambda_m)$, $m = 1000$ (the total number of simulations).

Step 3. In one simulation, suppose $n = 1000$ firms, $t = 500$ time periods, let the number of defaults ≤ 100 until time period $t = 500$, i.e default rate $\leq 10\%$. For the 4-state Markov chain model, at each time period, the credit rating dynamics can be established as a 4×4 matrix. Given an initial random number of firms in each state i, satisfying the total number of firms 1000, in each row of the matrix, the credit rating data are multinomial random numbers with the probabilities of the according row in the transition matrix P which is calculated in Step 2.

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\sum_{j=1}^{4} P_{ij} = 1$.

For time period t in $(0 \sim 1)$, we can get $data_1$ as

$$data_1 = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ 0 & 0 & 0 & d_{44} \end{pmatrix}$$

where $d_i = \sum_{j=1}^{4} d_{ij}$,$i \in K$, $K = 4$, and $d_1 + d_2 + d_3 + d_4 = 1000$ (the total number of firms), and for row i, d_{ij} follows multinomial random numbers with initial d_i, and $P_{i1}, P_{i2}, P_{i3}, P_{i4}$.

For time period t in $(1 \sim 2)$, we use the same method to get $data_2$, but the multinomial random numbers are generated with initial d_j, and $P_{i1}, P_{i2}, P_{i3}, P_{i4}$, where $d_j = \sum_{i=1}^{4} d_{ij}$ in $data_1$.

Follow this method of generating data, we can get $data_3$ with changing d_j in $data_2$, and so on, until we get $data_{500}$.

For 1000 simulations, redo step 2 for 1000 times using the same prior G_0 to get $\Lambda_1, \ldots, \Lambda_m$, $m = 1000$, and follow step 3 to generate time series data with length 500 under the same P_m, and change P_m in different simulations, $m = 1 \sim 1000$.

Step 4. Critical value determined for rule 1: Shewhart control chart detection method.

From equation(2.1.2), we can get $M(p)$ combining with step 1 and step 2 for $K = 4$. For 1000 simulations, we can get

$$\Lambda_1 \to P_1 \to M(p)_1$$

$$\vdots$$

$$\Lambda_m \to P_m \to M(p)_m, m = 1000$$

So the 95% quantile of the histogram for $M(p)$ can be seen as a critical value.

Step 5. Critical value determined for rule 2: Generalized likelihood ratio detection rule for a single change point model.

Using the data from step 3, at $T = 500$

$$V_2 = \max_{t_0 \le k \le T} \left\{ \frac{1}{T} \{ \sup_{\theta} \sum_{i=1}^{k} \log f_{\theta}(\boldsymbol{Y}_i) + \sup_{\tilde{\theta}} \sum_{i=k+1}^{n} \log f_{\tilde{\theta}}(\boldsymbol{Y_i}) - \sup_{\lambda} \sum_{i=1}^{n} \log f_{\lambda}(\boldsymbol{Y}_i) \} \right\} \tag{3.2.1}$$

From 1000 simulations, we can get 1000 V_2, and the 95% quantile of the histogram for V_2 can be seen as a critical value for rule 2.

Step 6. Critical value determined for rule 3: Detection rule of an extension of Shiryaev's Bayesian single change-point model.

Using the data from step 3, at $T = 500$

$$V_3 = \frac{1}{k_p} \left\{ \sum_{r=T-k_p}^{n} \frac{f_{1,r-1} f_{r,T}}{(1-p)^{T-r+1} f_{0,0} f_{1,T}} \ge \gamma_p \right\} \tag{3.2.2}$$

Like step 5, from 1000 simulations, we can get 1000 V_3, and the 95% quantile of the histogram for V_3 can be seen as a critical value for rule 3.

Step 7. Critical value determined for rule 4: Sequential detection rule for multiple structural breaks.

Using the data from step 3, at $T = 500$

$$V_4 = \frac{1}{k_p}\left\{ \sum_{r=T-kp}^{T} p_{r,T} \geq \eta_p \right\} \tag{3.2.3}$$

From 1000 simulations, we can get 1000 V_4, and the 95% quantile of the histogram for V_4 can be seen as a critical value for rule 4.

3.2.2 Simulation Results with Critical Value

In this study, we give four groups of prior to test the detection of single change-point model using the critical value,and each group is based on 1000 simulations. The data is generated by setting G_0 as the initial prior and G_1 as the changing prior following the steps from step 1 to step 3 in section(3.2.1) with the prior G_0 changed to G_1 at t=200. In the last part of this section, we give a discussion about the multiple change points model with critical values, in which the data is generated by changing the prior G_1 to G_2 at t=300 in addition to single change-point model. We choose $p = 0.01$, $n_p = 20$ and $k_p = 20$ in the detection rule of the extension of

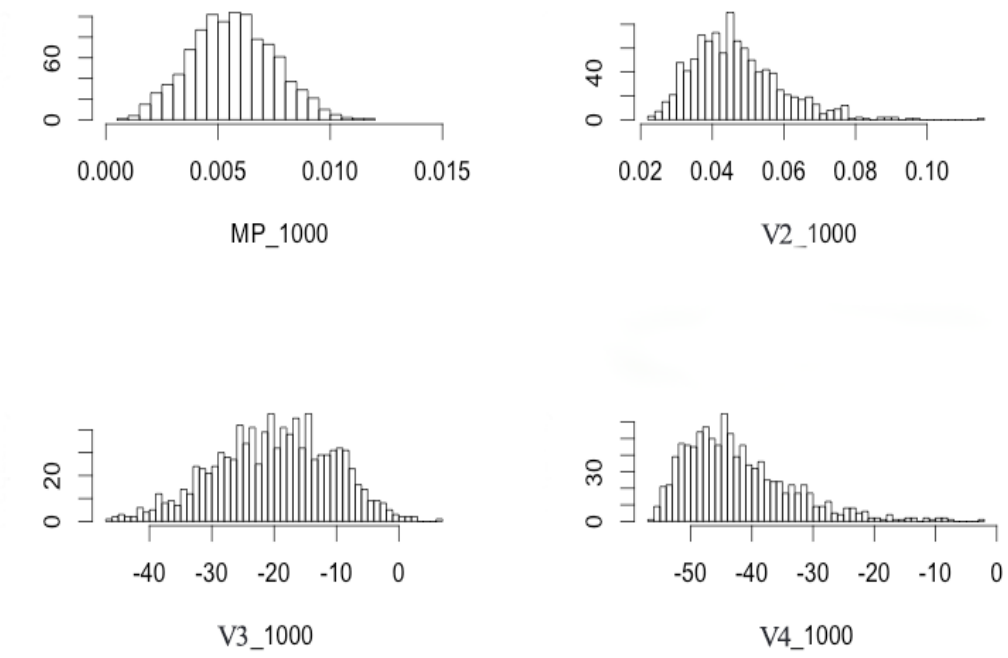

Figure 3.5: Histogram for $M(p)$ in rule 1, V_2 in rule 2, V_3 in rule 3 and V_4 in rule 4

Shiryaev's Bayesian change-point model (2.3.17) and the sequential detection rule for multiple structural breaks (2.4.9) and once a change-point is detected at time t, the data before t is not considered for the the next time detection, that means the detection process will entirely move on a window size of 20 to $t + 20$.

We also use some tables to compare the performance of the four rules. In each table, the false alarm rate is calculated as the number of pre-detection over the number of total simulations of 1000. The expected delay is $E[\tau - t_0 | \tau \leq 500]$, where τ is the delayed detected time, and t_0 is the true change-point time. All simulations

are considered based on the thresholds for 95% quantile, 97.5% quantile and 99% quantile of the histograms in 3.5, and the critical values are calculated using the method in section (3.2.1) from step 4 to step 7.

Group 1. Let the change-point time at t=200, given G_0:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.6 & 0 & 0.4 & 0.002 \\ 0.8 & 0.3 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

and G_1:

$$\alpha = \begin{pmatrix} 0 & 0.8 & 0.9 & 0.01 \\ 0.7 & 0 & 0.5 & 0.002 \\ 0.8 & 0.6 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 2 & 1 & 3 & 0 \end{pmatrix}$$

Table 3.4: False alarm rate and expected detection delay given 95%, 97.5%, 99% quantile for the change-point time at $t = 200$ under group 1

		95%	97.5%	99%
Rule (1)	False alarm rate	10.3%	8.8%	9.7%
	Expected delay	36.1	37.8	37.1
Rule (2)	False alarm rate	36.2%	20.1%	10.2%
	Expected delay	1.76	1.72	1.75
Rule (3)	False alarm rate	4.7%	2.1%	0.4%
	Expected delay	1.6	1.83	1.79
Rule (4)	False alarm rate	4.4%	1.8%	4%
	Expected delay	1.6	1.83	1.79

Group 2. Let the change-point time at t=200, given G_0:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.6 & 0 & 0.4 & 0.002 \\ 0.8 & 0.3 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

and G_1:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.7 & 0 & 0.5 & 0.002 \\ 0.8 & 0.6 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 2 & 1 & 1 & 0 \end{pmatrix}$$

Table 3.5: False alarm rate and expected detection delay given 95%, 97.5%, 99% quantile for the change-point time at $t = 200$ under group 2

		95%	97.5%	99%
Rule (1)	False alarm rate	10.5%	8.7%	10%
	Expected delay	39.1	40.8	40.2
Rule (2)	False alarm rate	37.7%	21.4%	12.5%
	Expected delay	1.85	1.91	1.92
Rule (3)	False alarm rate	4.9%	2.2%	1.3%
	Expected delay	1.8	2.09	2.12
Rule (4)	False alarm rate	4.4%	1.9%	1.2%
	Expected delay	1.8	2.09	2.12

Group 3. Let change-point time at t=200, given G_0:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.6 & 0 & 0.4 & 0.002 \\ 0.8 & 0.3 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

and G_1:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.7 & 0 & 0.5 & 0.002 \\ 0.8 & 0.6 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

Table 3.6: False alarm rate and expected detection delay given 95%, 97.5%, 99% quantile for the change-point time at $t = 200$ under group 3

		95%	97.5%	99%
Rule (1)	False alarm rate	10.3%	9.1%	10.1%
	Expected delay	39.1	40.8	40.2
Rule (2)	False alarm rate	36.5%	21%	11.1%
	Expected delay	1.66	1.63	1.62
Rule (3)	False alarm rate	5.5%	2.7%	1%
	Expected delay	1.6	1.61	1.81
Rule (4)	False alarm rate	4.7%	2.5%	1%
	Expected delay	1.64	1.61	1.8

Group 4. Let change-point time at t=200, given G_0:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.7 & 0 & 0.5 & 0.002 \\ 0.8 & 0.6 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

and G_1:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.6 & 0 & 0.4 & 0.002 \\ 0.8 & 0.3 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

Table 3.7: False alarm rate and expected detection delay given 95%, 97.5%, 99% quantile for the change-point time at $t = 200$ under group 4

		95%	97.5%	99%
Rule (1)	False alarm rate	4.1%	0.8%	3.3%
	Expected delay	35.9	44.3	38.1
Rule (2)	False alarm rate	32.4%	13.3%	4.8%
	Expected delay	1.33	1.17	1.26
Rule (3)	False alarm rate	3.7%	2%	0.6%
	Expected delay	1.12	1.19	1.37
Rule (4)	False alarm rate	2.5%	1.1%	0.3%
	Expected delay	1.11	1.21	1.37

From the table 3.4, table 3.5, table 3.6, table 3.7, which are the results for group 1 to group 4 respectively, both rule 3 and rule 4 give acceptable pre-detection rate and low expected delay, which are within 2 months, while the results from rule 1 and rule 2 either have very high pre-detection rate or high expected detection delay. There is no missed change-point detected in each simulation of all the four rules.

Now we move on to the multiple change points model. Group 5 lists the priors we have used to generate the data in this model and table 3.8 gives the result given 95% quantile, 97.5% quantile and 99%quantile thresholds.

5. Two change-points model with first change-point time at t_1=200, second change-point at t_2=300. Given G_0:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.6 & 0 & 0.4 & 0.002 \\ 0.8 & 0.3 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

and G_1:

$$\alpha = \begin{pmatrix} 0 & 0.8 & 0.9 & 0.01 \\ 0.7 & 0 & 0.5 & 0.002 \\ 0.8 & 0.6 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 2 & 1 & 3 & 0 \end{pmatrix}$$

and G_2:

$$\alpha = \begin{pmatrix} 0 & 0.2 & 0.5 & 0.01 \\ 0.7 & 0 & 0.5 & 0.002 \\ 0.8 & 0.6 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$$

Table 3.8: False alarm rate and expected detection delay given 95%, 97.5%, 99% quantile for two change points time at $t_1 = 200$, $t_2 = 300$ under group 5

	t_1=200			t_2=300		
Rule (1)	95%	97.5%	99%	95%	97.5%	99%
Detected rate	81.8%	80.1%	80.7%	99.9%	99.9%	99.8%
False alarm rate	19.9%	16.9%	18.6%	27%	19.6%	18.1%
Expected delay	20.97	21.15	21.01	18.87	37.02	48.11
Accurate rate	3.0%	1.6%	1.1%	1.3%	0.4%	0.3%
Rule (2)	95%	97.5%	99%	95%	97.5%	99%
Detected rate	100%	100%	100%	100%	100%	100%
False alarm rate	34.8%	20.5%	10.3%	19.4%	12.5%	5.3%
Expected delay	1.89	1.90	1.78	1.60	1.50	1.38
Accurate rate	62.5%	76.4%	86.0%	80.1%	86.9%	93.9%
Rule (3)	95%	97.5%	99%	95%	97.5%	99%
Detected rate	100%	100%	100%	100%	100%	100%
False alarm rate	4.7%	2.3%	1.1%	1.3%	0.6%	0.4%
Expected delay	2.083	2.077	2.067	1.67	1.60	1.56
Accurate rate	94.1%	96.4%	97.4%	97.5%	97.9%	98.0%
Rule (4)	95%	97.5%	99%	95%	97.5%	99%
Detected rate	100%	100%	100%	100%	100%	100%
False alarm rate	3.2%	1.4%	0.7%	0.5%	.3%	0.3%
Expected delay	2.0	2.08	2.0	1.64	1.60	1.56
Accurate rate	95.5%	97.3%	97.3%	98.1%	98.2%	98.1%

Table 3.8 is the results of the four types of detection rule based on the two change points model. In this model, we consider another two indexes: the detected rate and the accurate rate. The detected rate represents the number of simulations that we can detect the change-point among 1000 simulations. Here, we pre-specify all the time points we can detect before $t = 250$ as the first change-point detection result, while all the time-points we detect after $t = 250$ as the second change-point detection result. From this pre-specification, if there is no time-point detected before $t = 250$, implies there is no detected result for the first change-point, which implies a zero

detected rate for the first change-point. For the accurate rate, we define the rate as the number of correct detection result which just at the change-point time divided by the total number of simulations number of 1000. The false alarm rate and the expected delay are calculated as the same way with the single change-point model.

Now comparing the detection results based on the critical value with the detection results from section 3.1, both of the results show that rule 3 and rule 4 are more acceptable than rule 1 and rule 2 for the credit rating dynamics, and it has obvious lower expected delay and more accurate detected rate in section 3.2 with the theoretical thresholds.

Chapter 4

Real Data Analysis with Different Method of Threshold Determined

4.1 Data Description

In this study, the data set is obtained from COMPUSTAT and consists of Standard & Poor's monthly credit ratings from January 1985 to September 2009,which is consistent with the data used in Xing et al.(2012). The data set contains 21,755 firms with a total of 2,160,809 rating records, which were recorded at the end of each month. There are 10 rating categories, AAA, AA, A, BBB, BB, B, CCC, CC, C and D(default), and 25 subcategories. According to Xing et al.(2012), the data are

cleaned as follows, we first put both group C and group CC into group CCC and also remove the rating records of two invalid ratings "N.M.". Then we have eight rating categories, AAA, AA, A, BBB, BB, B, CCC, D. Note the first transition happens after it transfers away from its initial rating. We can see that there is only one rating transition among 1286 initial ratings in 1985. So our analysis is based on the data set from January 1986 to September 2009. Also, we should note that the starting time of entering our recorded database is different for different firms. Figure 4.1 gives a clear explanation of this for 3 firms.

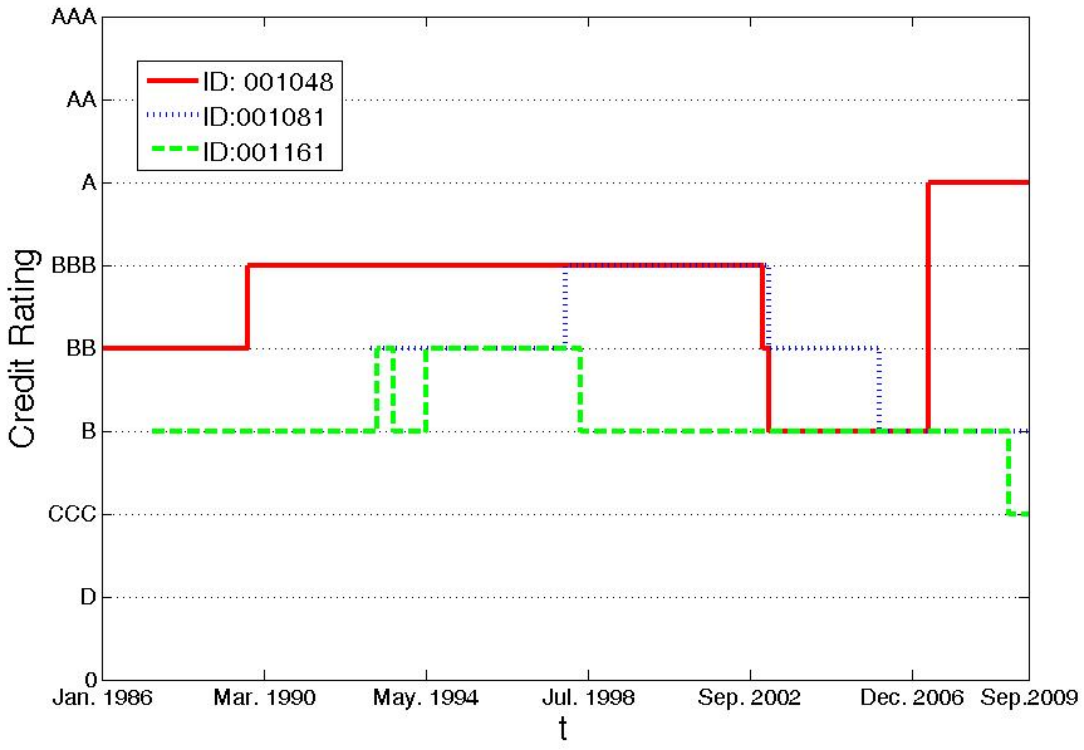

Figure 4.1: Credit Rating Recorded Data for 3 Firms

4.2 Parameters Estimation

4.2.1 EM Algorithm in Estimating Parameters of the Transition Matrix

In our model of the credit rating transition matrix, the parameter vector $\Phi = \{p, \alpha_{i,j}, \beta_i | (i,j) \in \mathcal{K}\}$, where $\mathcal{K} = \{(i,j) | i \neq j, 1 \leq i \leq K-1, 1 \leq j \leq K\}$ and for the 8 transition states in the real data analysis, $K = 8$. It is computationally extensive to maximize the likelihood function directly with the $K(K-1)+1$-dimemsional vector in Φ. In our study, we use the EM algorithm to complete the estimation for the parameters in this model.

The EM (expectation-maximization) algorithm is a general iterative algorithm for parameter estimation by maximum likelihood when the model depends on unobserved latent variables. The term EM was introduced in Dempster, Laird, and Rubin (1977) where the general results about the behavior of the algorithm was proved as well as many applications. The iteration in the EM algorithm alternates between the E-step, which is the function for the expectation of the log-likelihood based on the current estimation for the parameters, and an M-step, which calculates the parameters by maximizing the expected log-likelihood function from the E-step. The E-step and the M-step are repeated alternately until the difference between the likelihood function on two continuous time period $L(\theta_{t+1}) - L(\theta_t)$ is less than a prescribed small quantity δ.

Apply the EM Algorithm in this study with the 8-state transition matrix, the likelihood function is given in the Appendix B. in Xing et al.(2012) as follows:

$$
\begin{aligned}
l_c(\Phi) &= \sum_{l=1}^{L}\sum_{i=1}^{K}\Big\{\sum_{j\neq i} K_{t_{l-1},t_l}^{(i,j)} log\lambda_{t_{l-1},t_l}^{(i,j)} - \Big(\sum_{j\neq i}\lambda_{t_{l-1},t_l}^{(i,j)} + 1 - K\Big)S_{t_{l-1},t_l}^{(i)}\Big\} \\
&+ \sum_{l=1}^{L}\sum_{i=1}^{K}\Big\{\sum_{j\neq i}\Big(\alpha_{ij}-1\Big)log\lambda_{t_{l-1},t_l}^{(i,j)} - \Big(\sum_{j\neq i}\lambda_{t_{l-1},t_l}^{(i,j)}\Big)\beta_i \\
&+ \sum_{j\neq i}\Big(\alpha_{ij}log\beta_i - log\Gamma(\alpha_{ij})\Big)\Big\}\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}} \\
&+ \sum_{l=1}^{L}\Big\{\Big[log(1-p)\Big]\mathbb{1}_{\{\Lambda_{t_l}=\Lambda_{t_{l-1}}\}} + \Big(logp\Big)\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}\Big\}
\end{aligned}
\tag{4.2.1}
$$

Then, combine the E-step and M-step together to the parameter p, α_{ij} and β_i respectively.

The details of the maximization process on the E-step can be described as follows:

1. For the parameter p, the steps of the derivative to p of the likelihood function 4.2.1 based on E-step, let

$$
\sum_{l=1}^{L}\frac{1}{p-1}\mathbb{E}\Big[\mathbb{1}_{\{\Lambda_{t_l}=\Lambda_{t_{l-1}}\}}\Big] + \sum_{l=1}^{L}\frac{1}{p}\mathbb{E}\Big[\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}\Big] = 0
$$

we have

$$\sum_{l=1}^{L} \frac{1}{1-p}\Big[1-\mathbb{P}\Big(\Lambda_{t_l} \neq \Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)\Big] = \sum_{l=1}^{L} \frac{1}{p}\mathbb{P}\Big(\Lambda_{t_l} \neq \Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)$$

where $\mathbb{P}\Big(\Lambda_{t_l} \neq \Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big) = \mathbb{P}\Big(\mathbb{1}_l = 1|\mathcal{Y}_{(0,T)}\Big)$

then we can get the estimation of parameter p is

$$\hat{p} = \sum_{l=1}^{L} \mathbb{P}\Big(\Lambda_{t_l} \neq \Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)/L \tag{4.2.2}$$

2. For the parameter β_i, same as p, let

$$-\sum_{l=1}^{L}\sum_{i=1}^{K}\Big\{\mathbb{E}\Big[\sum_{j\neq i}\lambda^{(i,j)}_{t_{l-1},t_l}\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}\Big] + \sum_{j\neq i}\frac{\alpha_{ij}}{\beta i}\mathbb{E}\Big[\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}\Big]\Big\} = 0$$

then we have the estimation of β_i is

$$\hat{\beta}_{i,new} = \frac{\sum_{l=1}^{L}\sum_{j\neq i}\hat{\alpha}_{ij,old}\mathbb{P}\Big(\Lambda_{t_l} \neq \Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)}{\sum_{l=1}^{L}\mathbb{E}\Big[\Big(\sum_{j\neq i}\lambda^{(i,j)}_{t_{l-1},t_l}\Big)\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}|\mathcal{Y}_{(0,T)}\Big]} \tag{4.2.3}$$

3. For the parameter α_{ij}, the estimation is not obvious. First we have

$$\sum_{l=1}^{L}\sum_{i=1}^{K}\left\{\mathbb{E}\Big[log\lambda^{(i,j)}_{t_{l-1},t_l}\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}\Big]+\sum_{j\neq i}log\beta_i\mathbb{P}\Big(\Lambda_{t_l}\neq\Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)\right.$$
$$\left.-\frac{\Gamma'(\alpha_{ij})}{\Gamma(\alpha_{ij})}\mathbb{P}\Big(\Lambda_{t_l}\neq\Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)\right\}=0 \tag{4.2.4}$$

From the above equation 4.2.4, we can get

$$\frac{\Gamma'(\hat{\alpha}_{ij,new})}{\Gamma(\hat{\alpha}_{ij})}=\frac{\sum_{l=1}\mathbb{E}\Big(log\lambda^{(i,j)}_{t_{l-1},t_l}\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}|\mathcal{Y}_{(0,T)}\Big)}{\sum_{l=1}^{L}\mathbb{P}\Big(\Lambda_{t_l}\neq\Lambda_{t_{l-1}}|\mathcal{Y}_{(0,T)}\Big)}+log\hat{\beta}_{i,old} \tag{4.2.5}$$

and from Xing et al.2012, it was showed that

$$\mathbb{E}\Big(\lambda^{(i,j)}_{t_{l-1},t_l}\mathbb{1}_{\{\Lambda_{t_l}\neq\Lambda_{t_{l-1}}\}}|\mathcal{Y}_{(0,T)}\Big)=\sum_{l\leq k\leq L}\pi_{lkk}\frac{K^{(i,j)}_{t_l,t_k}+\alpha_{ij}}{S^{(i)}_{t_l,t_k}+\beta_i}$$

Then the equation 4.2.5 can be solved numerically by grid search with digamma function

$$Diga(\frac{r}{M})=-\gamma-\ln(2M)-\frac{\pi}{2}\cot(\frac{r\pi}{M})+2\sum_{i=1}^{\frac{M}{2}}\cos(\frac{2\pi ir}{M})\ln\sin(\frac{\pi i}{M})$$

4.2.2 Initial Value Determined

In this study, we use the sample moments to estimate the initial α_{ij} and βi.

1. For any time period $(t_{l-1}, t_l))$, we let

$$\begin{cases} \lambda^{(i,j)} = \frac{K^{(i,j)}}{S^{(i)}} \\ \lambda^{(i,i)} = -\sum_j \frac{K^{(i,j)}}{S^{(i)}} \end{cases}$$

then the expectation of λ and λ^2 in the time period from $(0 \sim L)$ is

$$\begin{cases} m_1 = \mathbf{E}(\lambda^{(i,j)}) = \sum_{t=1}^{L} \lambda_t^{(i,j)}/t \\ m_2 = \mathbf{E}((\lambda^{(i,j)})^2) = \sum_{t=1}^{L} (\lambda^{(i,j)})^2/t \end{cases}$$

if $m_2 \neq m_1^2$, let $\alpha_{ii} = 0$, $\beta_{ii} = 0$

$$\alpha_{ij} = \frac{m_1^2}{m_2 - m_1^2} = \frac{\mathbf{E}^2(\lambda^{(i,j)})}{\mathbf{E}((\lambda^{(i,j)})^2) - \mathbf{E}^2(\lambda^{(i,j)})} \tag{4.2.6}$$

$$\beta_{ij} = \frac{m_1}{m_2 - m_1^2} = \frac{\mathbf{E}(\lambda^{(i,j)}}{\mathbf{E}((\lambda^{(i,j)})^2) - \mathbf{E}^2(\lambda^{(i,j)})} \tag{4.2.7}$$

and

$$\beta_i = \sum_{j=1}^{K} \beta_{i,j}/(N(\beta_{i,j})) \tag{4.2.8}$$

where $N(\beta_{i,j})$ is the number of non-zero $\beta_{i,j}$ for specific i.

4.3 Threshold Determined from the Simulations Data

From the description in section 4.1, our data set has totally L=285 time periods from January 1986 to September 2009. To determine the threshold for the detection in the real data analysis, we first put the data set into two parts, and follow the EM algorithm in section 4.2.1 to estimate all the parameters $\Phi = \{p, \alpha_{i,j}, \beta_i | (i,j) \in \mathcal{K}\}$ in the 8-state Markov chain model using the first part of the data set, then simulate data under the estimated prior to determine the threshold with $95\%, 97.5\%$ or 99%, last, apply the four detection methods to the second part of the data set to have an early detect of the credit market. The detailed process is stated as follows.

To make the estimation more accurate, in this study, we start the detection from $t = 201$, which means the estimation is based on at least 200 monthly data. For the time point $t = 201$, we first follow the estimation steps in 4.2.1 based on the data set from $t = 1$ to $t = 200$ to get an estimation prior of $G_{t=200}$. Given the prior $G_{t=200}$, the simulation data is generated under the method in section 3.2.1. With 1000 simulation data, we have a 95% quantile, 97.5% quantile and 99% quantile

critical value for all four kinds of rules under the function $M(p)$ in (2.1.2), V_2 in (3.2.1), V_3 in (3.2.2) and V_4 in (3.2.3). Then we respectively detect the real data under rule 1, rule 2, rule 3 and rule 4 for the time period $(1 \sim 201)$, and check if there is a change-point at $t = 201$.

Now we move on the real data set to $t = 202$ and first have an estimated prior $G_{t=201}$ and simulate 1000 groups of data under $G_{t=201}$. Following the same steps as $t = 200$, calculate the critical value for all four kinds of rules and detect the real data from the time period $(1 \sim 202)$ and check if there is a change-point at $t = 202$.

Repeat the process for $t = 203$, $t = 204$, $\ldots$, $t = 285$. If there is a change-point detected at time t, the data before t is not considered for the next time detection. This step eliminates the continued impact from a structure break happened. In rule 4 (the sequential detection rule for multiple structure breaks), the detection process entirely move on a window size to $t + kp$, and $kp = 20$ in this study. And the window size kp should not be set too large to cover all practical data for detection.

We also use a table to list the result for this study. The detection time perids in our study is actually from $t = 201$, which according to Sep, 2002. From the table 4.1, synthesize the result from all four kinds of detection rules, especially based on the result from rule 3 and rule 4, in which we have gotten more acceptable results from simulations in chapter 3, we can summarize that a structure break during 2002 and 2003, which can be seen as a detection delay for the second recession of March 2001

- Nov 2001 announced by NBER. A change-point series among Mar, 2007 to Aug, 2009 can be seen from the result which matches the third recession of Dec, 2007 to June, 2009 announced by NBER.

Table 4.1: Structure break detected based on the data from Jan 1986 to Sep 2009

	Structure break				
Rule (1)	226 (Oct,2004)	227 (Nov,2004)	228 (Dec,2004)	230 (Feb,2005)	231 (Mar,2005)
	233 (May,2005)	236 (Aug,2005)	245 (May,2006)	255 (Mar,2007)	276 (Dec,2008)
	278 (Feb,2009)				
Rule (2)	210 (Jun,2003)	213 (Sep,2003)	214 (Oct,2003)	225 (Sep,2004)	238 (Oct,2005)
	255 (Mar,2007)	275 (Dec,2008)	276 (Aug,2009)	284 (Nov,2008)	
Rule (3)	206 (Feb,2003)	222 (Jun,2004)	227 (Nov,2004)	238 (Oct,2005)	240 (Dec,2005)
	252 (Dec,2006)	257 (May,2007)	259 (Jul,2007)	260 (Aug,2007)	261 (Sep,2007)
	275 (Nov,2008)	276 (Dec,2008)	283 (Jul,2009)	284 (Aug,2009)	
Rule (4)	218 (Feb,2004)	276 (Dec,2008)			

4.4 Threshold Determined under Bootstrapping Method

In this part, we determine the threshold using bootstrapping method to resample 1000 groups of time series data from January 1986 to September 2009 which is based

on the data set we have described in section 4.1. The method for the critical value determined is very similar as the steps stated in section 4.3, the difference is that we do not simulate data from the estimated priors, instead we use 1000 groups of bootstrap sample data set to determine the threshold for each time point from $t = 201$, $t = 202$, ..., $t = 285$.

Step 1, suppose we have N firms with transition history from $t = 1$ to $t = 285$, for bootstrap sample 1, we resample $(1, 2, \ldots, N)$ to get a new all-firms transition history. Redo bootstrap for sample 2, 3, ... B=1000 to have a total of 1000 groups of time series data.

Step 2, for $t = 201$, following the EM Algorithm estimation steps in 4.2.1 and 4.2.2 based on the real data set from $t = 1$ to $t = 201$, we then get an estimation prior of $G_{t=201}$ and an estimation of parameter p.

Step 3, given the prior $G_{t=201}$ and p, we can calculate the value $M(p)$ in (2.1.2), V_2 in (3.2.1), V_3 in (3.2.2) and V_4 in (3.2.3) under specific kp based on the real data set for rule 2, rule 3 and rule 4. In this study, we set $kp = 15$ as the window size for detection rule 2 and $kp = 20$ as the window size for detection rule 3 and rule 4 in this study.

Step 4, from the 1000 groups of the time series data which we have bootstrapped in step 1, we can get 1000 values of $M(p)$, V_2, V_3 and V_4 respectively for each rule.

Then a quantile for $M(p)$, V_2, V_3 and V_4 from the real data set in step 3 to the values from the 1000 bootstrapping data can be obtained for each rule respectively. This quantile can be seen as the p-value in this study. If this p-value is more than 90%, we will claim there is a change-point at this time $t = 201$, or if there is a big change in this quantile compared to the previous time point, we also claim there is a big change at this time in the credit rating transitions.

Step 5, repeat the step 2 to step 4 for $t = 202$, $t = 203$, ..., $t = 285$. As stated in section 4.3, once there is a change-point detected at time t, the data before t is not considered for the next time detection.

For the result of the bootstrapping method, we use a table to describe the change-point time that we have claimed under this method, in addition, all the quantiles (p-value) calculated for the four rules are listed in another table.

Table 4.2: Structure break detected based on the data from Jan 1986 to Sep 2009 under bootstrapping method

	Structure break		
Rule (1)	205 (Jan,2003)	212 $\sim$ 229 (Aug,2003)$\sim$(Jan,2005)	239 $\sim$ 242 (Nov,2005)$\sim$(Feb,2006)
	247 $\sim$ 253 (Jul,2006)$\sim$(Jan,2007)	261 (Sep,2007)	264 (Dec,2007)
	272 (Aug,2008)	283 (Jul,2009)	285 (Sep,2009)
Rule (2)	209 (May,2003)	212 (Aug,2003)	216 (Dec,2003)
	218 (Feb,2004)	238 (Oct,2005)	265 (Jan,2008)
	269 (May,2008)	270 (Jun,2008)	275 (Nov,2008)
	280 (Apr,2009)		
Rule (3)	203 (Nov,2002)	208 (Apr,2003)	216 (Dec,2003)
	226 (Oct,2004)	273 (Sep,2008)	276 (Dec,2008)
	277 (Jan,2009)	279 (Mar,2009)	
Rule (4)	220 (Apr,2004)	224 (Aug,2004)	227 (Nov,2004)
	235 (Jul,2005)	239 (Nov,2005)	261 (Sep,2007)
	275 (Nov,2008)	278 (Feb,2009)	284 (Aug,2009)

Table 4.3: Quantile (p-value) under bootstrapping method at time 201 $\sim$ 285

Month	T	R-1	R-2	R-3	R-4	Month	T	R-1	R-2	R-3	R-4
09/2002	201	67.1%	3.6%	50.6%	7.7%	01/2006	241	99.3%	3.8%	26.6%	0.8%
10/2002	202	41.7%	8.1%	54.7%	11.9%	02/2006	242	95.9%	2.7%	28.1%	0%
11/2002	203	87.0%	12.2%	100%	4%	03/2006	243	82.5%	2%	24.7%	0%
12/2002	204	21.6%	13.2%	47.9%	14.8%	04/2006	244	76%	1.2%	27.3%	0.2%
01/2003	205	99.2%	11.6%	44.9%	16.3%	05/2006	245	40.4%	1.8%	23%	1.1%
02/2003	206	69.9%	9.9%	44.5%	18.7%	06/2006	246	28.8%	1%	18.4%	8%
03/2003	207	88.6%	6.1%	47.5%	23.1%	07/2006	247	98.2%	0.5%	14.9%	3.6%
04/2003	208	80.5%	4.5%	100%	34.4%	08/2006	248	92.9%	0.7%	18.6%	2.1%
05/2003	209	89.7%	10.6%	45.2%	8.4%	09/2006	249	98.7%	1.8%	19.6%	2.6%
06/2003	210	48.5%	4.9%	45.2%	9.7%	10/2006	250	70%	4%	23.1%	0%
07/2003	211	33.7%	6.1%	41.3%	22%	11/2006	251	95.1%	8.2%	26.3%	2.4%
08/2003	212	94.1%	17.2%	38.2%	19.4%	12/2006	252	87.2%	16%	27.5%	1.2%
09/2003	213	98.9%	18.9%	37.6%	4.1%	01/2007	253	97%	11.2%	28.3%	0.6%
10/2003	214	97.8%	25.9%	30.8%	20.9%	02/2007	254	17.5%	3.2%	31.8%	3.1%
11/2003	215	98.5%	29.4%	28.8%	30.4%	03/2007	255	13.2%	2.2%	33.4%	6.4%
12/2003	216	98.9%	34.1%	0.1%	44.1%	04/2007	256	18.6%	10.6%	30.4%	2.6%
01/2004	217	99.5%	27.3%	13.9%	40.3%	05/2007	257	13.4%	24.4%	30.1%	13.6%
02/2004	218	99.5%	36.4%	18.1%	51.4%	06/2007	258	58.9%	18.8%	36.9%	6.5%
03/2004	219	99.7%	21.3%	22.8%	56.3%	07/2007	259	68.9%	4.1%	39.3%	0.3%
04/2004	220	98.5%	19.9%	24.8%	96.5%	08/2007	260	56.7%	3.1%	37.9%	0.5%
05/2004	221	96.8%	18.6%	39.4%	57.1%	09/2007	261	98.2%	2%	35.7%	17.6%
06/2004	222	96.2%	18.7%	33.4%	59.5%	10/2007	262	71.4%	4.8%	38.5%	1.9%
07/2004	223	97.1%	21.4%	34.4%	57.5%	11/2007	263	68%	2.5%	40.5%	1.2%
08/2004	224	91.1%	11.1%	38.9%	90%	12/2007	264	91%	3%	37.7%	4%
09/2004	225	90%	11.3%	32.4%	41.1%	01/2008	265	48.7%	13.8%	37.7%	4.3%
10/2004	226	83.2%	4%	4.9%	12.5%	02/2008	266	58.1%	13.2%	32%	1.3%
11/2004	227	91.2%	2.6%	24.7%	82%	03/2008	267	33.4%	15.6%	35%	0%
12/2004	228	86%	3.1%	32.9%	53.1%	04/2008	268	26.5%	14.9%	34.6%	0.5%
01/2005	229	94.8%	2.7%	35.3%	58.7%	05/2008	269	66.6%	27.1%	34%	0.8%
02/2005	230	45.5%	0.9%	34.2%	37.6%	06/2008	270	78.3%	33.3%	31.9%	3%
03/2005	231	50.4%	1.1%	43.4%	27.4%	07/2008	271	16.2%	24.1%	25%	3.7%
04/2005	232	70.3%	1.8%	43.6%	35.5%	08/2008	272	97%	13%	22.5%	9.4%
05/2005	233	61.3%	4.2%	44.7%	37.3%	09/2008	273	70.3%	1.6%	10.3%	8.6%
06/2005	234	74.3%	3.1%	41.5%	33.3%	10/2008	274	3%	1.5%	9.5%	0.5%
07/2005	235	86.9%	2.8%	38.3%	5%	11/2008	275	6.3%	17.2%	9.7%	37.2%
08/2005	236	35.1%	2.1%	29%	26.7%	12/2008	276	0%	31.8%	3.5%	25.4%
09/2005	237	62.5%	4.1%	27.8%	1.7%	01/2009	277	0.4%	46.7%	39.7%	45.1%
10/2005	238	37.1%	14.6%	27.5%	24.1%	02/2009	278	3.4%	53.2%	49.5%	88.5%
11/2005	239	91.7%	2.3%	22.5%	0.2%	03/2009	279	0.2%	53.4%	60.2%	83.5%
12/2005	240	93.2%	0.4%	23%	0%	04/2009	280	4.1%	56.1%	62.8%	77.4%
05/2009	281	64.4%	53.6%	63.9%	82.1%	08/2009	284	87.8%	42.3%	59.8%	92.1%
06/2009	282	32.3%	53.2%	63.4%	78%	09/2009	285	100%	21.2%	61.7%	88.5%
07/2009	283	99.6%	42%	62.9%	89.8%						

4.5 Concluding Remarks

Table 4.1 and 4.2 respectively list the change-point time we have claimed under different threshold determined methods. Compared the two tables, it is clearly to see a change-point during Jan, 2003 (205) $\sim$ Aug, 2003 (212), a change-point around Oct, 2005 (238), a change-point around Jan, 2008 (265) and a change-point during Nov, 2008 (275) $\sim$ Sep, 2009 (285). Combining with the economic recessions announced by NBER, the time periods detected in this study cover one year delayed reaction of the second recession (March, 2001-November, 2003), and the third recession (December 2007-June, 2009). The change-point we claimed during Jan, 2003 (205) $\sim$ Aug, 2003 (212) can been seen as a delayed credit rating reaction to the second economic recession, the change-point claimed around Jan, 2008 (265) is just the beginning of the third economic recession, and the change-point during Nov, 2008 (275) $\sim$ Sep, 2009 (285) is consistent with the cause of financial crisis staring in 2007. As known from the simulation study, there usually exists detection delay or false alarm. The results between the two methods are not exactly matched can been seen as the detection delay or false alarm problem in the real data study. That's another reason for us to use different threshold determined methods in this study and combine the results to the analysis of the real world.

Part II

Multivariate Log-linear Poisson Autoregression

Chapter 5

Introduction

5.1 Literature Review

Developing models for time series of counts has been a renewed interest with the widely applications in economics, finance, epidemiology and so on. Cox (1981) classifies the time series of counts into two categories: parameter-driven models and observation-driven models. From the recent research study on time series of counts, the observation driven models are very important and realistic (Kedem and Fokianos (2002), Tjøstheim (2012)). Many research are based on the observation driven models that assume the observations following a Poisson distribution, such as Poisson integer-valued GARCH (INGARCH) in Ferland et al.(2006), Fokianos et al.(2009)

and Zhu (2012); log-linear Poisson autoregression models in see Fokianos and Tjøstheim (2011); Poisson threshold models in Henderson et al.(2011); other observation-driven models for Poisson time series of counts in Davis et al.(2003), Davis and Liu (2012) and Neumann (2011). In this thesis, we also focus on the observation driven models that assume the observations following a Poisson distribution.

Fokianos et al.(2009) considers the geometric ergodicity and likelihood-based inference for Poisson autoregression in linear and nonlinear models and has shown that the differences between the perturbed and nonperturbed models vanish as far as considering the asymptotic distribution of the parameter estimations. In many industries like finance, biology, and epidemiology, multivariate time series of counts data are considered for analysis. So a bivariate Poisson autogression is proposed and the stability is proved in Liu (2012) chapter 4. Only positive association and the exclusive covariates which guarantee the positive regression term are two shortcomings for the Poisson autoregression model. To solve the drawbacks of the Poisson autoregression model, a univariate log-linear Poisson autoregression model in Fokianos and Tjøstheim (2011) is proposed with both negative and positive association taken into account and a straightforward time dependent covariates are accommodated. Our study in this part is to generalize the univariate log-linear Poisson autoregression into the multivariate case to investigate the interdependence between components of a vector time series of counts., which is an extension of log-linear Poisson autoregression model in Fokianos and Tjøstheim (2011).

5.2 Parameter-Driven Models for Poisson Counts

Let Y_1, ..., Y_n be a Poisson time series of counts, and denote x_t as the explanatory regression vector respect to Y_t at time t. Assume $Y_t|\lambda_t$ is Poisson(λ_t). The distribution of the Y_t given x_t and a stochastic process α_t are independent Poisson distributed with mean

$$\lambda_t = exp\{x_t^T\beta + Z_t\} \tag{5.2.1}$$

and

$$log\lambda_t = x_t^T\beta + Z_t \tag{5.2.2}$$

where $\{Z_t\}$ is a stationary Gaussian process and β is the regression coefficients vector. A simple example for Z_t is AR(1) process:

$$(Z_t + \sigma^2/2) = \phi(Z_{t-1} + \sigma^2/2) + \varepsilon_t \tag{5.2.3}$$

and $\{\varepsilon_t\} \sim$ IID N(0, $\sigma^2(1-\phi^2)$. The stability properties of the models are easy to derive and the regression parameters can be interpreted, $E(Y_t) = exp(x_t^T\beta)E[exp(Z_t)] = exp(x_t^T\beta)$, when $E[exp(Z_t)]$=1. (see Davis, R., Rodriguez-Yam, G.,). However, the estimation of this model is difficult because the likelihood function which has multiple integrals with sample size n is difficult to calculate. And the predictions based on this model are difficult.

5.3 Observation-Driven Models for Poisson Counts

Following the same assumption in equation 5.2.1, assume $Y_t|\lambda_t$ is Poisson(λ_t), rewrite equation 5.2.2

$$log\lambda_t = x_t^T\beta + Z_t \tag{5.3.1}$$

where $\{Z_t\}$ is a function of past observations Y_s , $s < t$. For example, $Z_t = a_1Y_{t-1} + a_2Y_{t-2} + ... + a_hY_{t-h}$. Unlike the parameter-driven model, the conditional mean λ_t in the observation-driven model solely depends on the past observations, which make the estimation easier and the prediction straightforward. However, the stability such as stationarity and ergodicity of the process is difficult to derive. In addition, $x_t^T\beta$ is not easy to interpret. Take the above example of Z_t, $E(Y_t) = exp(x_t^T\beta)E[exp(a_1Y_{t-1} + ... + a_hY_{t-h}])$.

A well-known observation-driven model for Poisson counts which was first proposed by Davis et al.(2003) is the generalized linear ARMA (GLARMA) process for Poisson counts. Assume $Y_t|\lambda_t \sim$ Poisson(λ_t), with $log\lambda_t = x_t^T\beta + Z_t$, where $Z_t = \sum_{i=1}^{\infty}\psi_i e_{t-i}$ and $e_i = (Y_t - \lambda_t)/\lambda_t^{\alpha}$ for $\alpha > 0$, $\{e_i\}$ is the martingale difference sequence. It was shown by Davis et al.(2003) that when $1/2 \leq \alpha \leq 1$, the chain $\{log\lambda_t\}$ has a stationary distribution. As the observation-driven models, the stability of GLARMA should be under some restrictive conditions for parameter space.

5.4 Poisson Autoregression

The generalized autoregressive conditional heteroscedasticity (GARCH)-type Poisson time series model, also known as Poisson autoregression was proposed by Fokianos et al.(2009), which is a famous observation-driven model for Poisson counts. In this model, suppose $\{Y_t\}$ follows a Poisson time series of counts, let $\{\mathcal{F}_t^{Y,\lambda}\}$ be a σ-field generated by $\{Y_0, Y_1, ..., Y_t, \lambda_0\}$, where $\{\lambda_t\}$ is a Poisson intensity process. The conditional mean λ_t of Y_t is a linear autoregression model depends on the previous mean and previous observation:

$$\lambda_t = d + \alpha\lambda_{t-1} + \beta Y_{t-1} \tag{5.4.1}$$

To ensure the stability of the model, the parameters d, α, β should be positive and must satisfy $\alpha + \beta < 1$. (See Rydberg and Shephard (2000), Streett, S. (2000), Fokianos et al.(2009), Ferland et al.(2006), and Neumann (2011)).

5.5 Log-linear Poisson Autoregression

To solve the drawbacks of the only positive association and the exclusive covariates for positive regressive terms, Fokianos and Tjøstheim (2011) proposed a new model for liner Poisson autogression, which take logarithm of the conditional mean and the associated observations with a small adjustment. Let $\{v_t\} \equiv log\lambda_t$, the

log-linear auroregression model for $\{Y_t\}$ is:

$$v_t = d + \alpha v_{t-1} + \beta log(Y_{t-1} + 1) \tag{5.5.1}$$

The stationarity and ergodicity has been proved in Fokianos and Tjøstheim (2011) that under $|\alpha| < 1$, when $\beta > 0$, then $|\alpha+\beta| < 1$, and when $\beta < 0$, then $|\alpha||\alpha+\beta| < 1$. The reason using $log(Y_{t-1}+1)$ instead of $log(Y_{t-1})$ is that the stability of the system with $log(Y_{t-1})$ is guaranteed only when $b < 0$ (see Fokianos and Tjøstheim (2011)).

5.6 Outline

This dissertation research in this part focus on the multivariate Poisson time series of counts data. We first proposed a bivariate log-linear Poisson autoregression model based on the log-linear Poisson autoregression model in Fokianos and Tjøstheim (2011). A stationarity and ergodicity has been proved in chapter 6, and we derive the likelihood inference for this bivariate log-linear Poisson autoregression model. In addition, not only for bivariate model, we extend our work to multivariate model in the simulation study. In real data study, we apply this model to analyze the association between two or three stocks/financial markets simultaneously. This theory can be applied to the number of transactions of financial time series of counts for two or more stocks and the number of exceedance returns for two or more stocks/ financial markets. We further use this model to estimate the exceedance returns asso-

ciation between SP500 and NASDAQ monthly counts record from 1996 to 2015, and the exceedance returns association among stock market indices in the US, Europe, and Japan.

Chapter 6

Multivariate Log-linear Poisson Autoregression Model

6.1 Model Specification

Denote $\{Y_{t,1}, Y_{t,2}\}$ be bivariate observations, and $\{Y_{t,1}, t \geq 1\}$ and $\{Y_{t,2}, t \geq 1\}$ are two time series of counts. Let $\{\lambda_{t,1}\}$ and $\{\lambda_{t,2}\}$ be the Poisson intensity process for $\{Y_{t,1}\}$ and $\{Y_{t,2}\}$, respectively. The bivariate Poisson autoregression was proposed in

Liu (2012) as: $\mathbf{Y}_t|\mathcal{F}_{t-1} \sim BP(\lambda_{t_1}, \lambda_{t_2}, \phi)$,

$$\begin{pmatrix} \lambda_{t,1} \\ \lambda_{t,2} \end{pmatrix} = \boldsymbol{D} + \boldsymbol{A} \begin{pmatrix} \lambda_{t-1,1} \\ \lambda_{t-1,2} \end{pmatrix} + \boldsymbol{B} \begin{pmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{pmatrix} \tag{6.1.1}$$

where, $\boldsymbol{D} \in \mathbb{R}_+^2$, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are 2×2 matrix with nonnegative entries, and $Cov(Y_{t_1}, Y_{t_2}|\mathcal{F}_{t-1}) = \phi$.

To solve the drawback of only positive association in 6.1.1, following the univariate log-linear Poisson autoregression, we propose a K-dimensional log-linear Poisson autoregression model. For simplicity, we first let $\{Y_{t,1}, t \geq 1\}$, $\cdots$, $\{Y_{t,k}, t \geq 1\}$ be two conditional independent time series of counts given information up to time $t-1$, let $\{v_{t,1}\} \equiv log\lambda_{t,1}$, $\cdots$, $\{v_{t,k}\} \equiv log\lambda_{t,k}$, so the multivariate log-linear Poisson autoregression model is:

$$\begin{pmatrix} v_{t,1} \\ \vdots \\ v_{t,k} \end{pmatrix} = \boldsymbol{D} + \boldsymbol{A} \begin{pmatrix} v_{t-1,1} \\ \vdots \\ v_{t-1,k} \end{pmatrix} + \boldsymbol{B} \begin{pmatrix} log(Y_{t-1,1}+1) \\ \vdots \\ log(Y_{t-1,k}+1) \end{pmatrix} \tag{6.1.2}$$

where $\boldsymbol{D} \in \mathbb{R}_+^K$ and the entries of the parameter matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ are belong to $\mathbb{R}$ with restrictions in the parameter space. The model 6.1.2 captures the dependence among $\{v_{t,1}\} \cdots \{v_{t,k}\}$, which means it is suitable to the coefficients of $\boldsymbol{A}$ or $\boldsymbol{B}$ not diagonal.

Specifically, for K=2, a bivariate log-liner Poisson autoregression is defined as

$$\begin{pmatrix} v_{t,1} \\ v_{t,2} \end{pmatrix} = \boldsymbol{D} + \boldsymbol{A} \begin{pmatrix} v_{t-1,1} \\ v_{t-1,2} \end{pmatrix} + \boldsymbol{B} \begin{pmatrix} log(Y_{t-1,1}+1) \\ log(Y_{t-1,2}+1) \end{pmatrix} \tag{6.1.3}$$

Assume $\{v_{0,1}, v_{0,2}\}$ and $\{Y_{0,1}, Y_{0,2}\}$ are fixed, and $\boldsymbol{I}$ is a 2×2 identity matrix. By recursion, model 6.1.3 can be expressed as

$$\begin{pmatrix} v_{t,1} \\ v_{t,2} \end{pmatrix} = \frac{\boldsymbol{I} - \boldsymbol{A}^t}{\boldsymbol{I} - \boldsymbol{A}} \boldsymbol{D} + \boldsymbol{A}^t \begin{pmatrix} v_{0,1} \\ v_{0,2} \end{pmatrix} + \sum_{i=0}^{t-1} \boldsymbol{A}^i \boldsymbol{B} \begin{pmatrix} log(Y_{t-i-1,1}+1) \\ log(Y_{t-i-1,2}+1) \end{pmatrix} \tag{6.1.4}$$

where the largest absolute eigenvalue of $\boldsymbol{A}$ is less than 1. Since the unobserved process $\{v_{t,1}, v_{t,2}\}$ is determined by the matrix version of previous functions of lagged observations, model 6.1.3 is still belongs to the class of observation driven models as defined by Cox (1981), but a matrix version.

6.2 Stationarity and Ergodicity

Before giving the results of the stationarity and ergodicity, we first have some preliminaries which are related notations and definitions we will use in this part. For

the ease of discussion, we simply consider $K = 2$ in the proof of the properties and it is similar arguments and discussions for the multivariate case of $K \geq 3$.

First, some definitions for norm. The vector norm of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ is defined as $||\boldsymbol{x}||_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}} = \sqrt[p]{|x_1|^p + |x_2|^p + \cdots + |x_n|^p}$. The inequality of vector norm is $||\boldsymbol{x}||_r \leq n^{(\frac{1}{r} - \frac{1}{p})}||\boldsymbol{x}||_p$, where n is the dimension of vector $\boldsymbol{x}$, and this inequality can be proved by Cauchy-Schwarz inequality. For matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, the p-induced norm of matrix $\boldsymbol{A}$ is defined as

$$||\boldsymbol{A}||_p = \max_{\boldsymbol{x} \neq 0} \frac{||\boldsymbol{A}||_p}{||\boldsymbol{x}||_p} = \max_{\boldsymbol{x} \neq 0} \frac{(\sum_{i=1}^m |\sum_{j=1}^n a_{ij} x_j|^p)^{\frac{1}{p}}}{(\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}}$$

where $\boldsymbol{x}$ is an $n \times 1$ vector. For $p = 1$, $||A||_1 = \max\limits_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$, which is the maximum absolute column sum of $\boldsymbol{A}$; for $p = \infty$, $||A||_\infty = \max\limits_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$, which is the maximum absolute row sum of $\boldsymbol{A}$. When $\boldsymbol{A}$ is a square matrix, let $\rho(\boldsymbol{A})$ be the spectral radius of $\boldsymbol{A}$, i.e. the largest absolute eigenvalue of $\boldsymbol{A}$, and $\rho(\boldsymbol{A}) \leq ||\boldsymbol{A}||_p$, for $1 \leq p \leq \infty$. If $\boldsymbol{A}$ is diagonal, then $\rho(\boldsymbol{A}) = ||\boldsymbol{A}||_1 = ||\boldsymbol{A}||_\infty$. Also, $\lim\limits_{r \to \infty} ||\boldsymbol{A}^r||^{\frac{1}{r}} = \rho(\boldsymbol{A})$.

We also need some related notations on the bivariate Markov chain $\{\boldsymbol{v}_t, t \geq 1\}$, where $\boldsymbol{v}_t = (v_{t,1}, v_{t,2})^T$. Combining with the notation in Fokianos and Tjøstheim (2011), independent Poisson process of unit intensity is $\boldsymbol{Y}_t = N(\boldsymbol{\lambda}_t)$, and

$$\boldsymbol{v}_t = \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_{t-1} + \boldsymbol{B}log(\boldsymbol{Y}_t + \boldsymbol{e}) \tag{6.2.1}$$

where $\boldsymbol{Y}_t = (Y_{t,1}, Y_{t,2})^T$, $\boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^T$, $\boldsymbol{v}_t = (v_{t,1}, v_{t,2})^T$, $\boldsymbol{e} = (1, 1)^T$, $\boldsymbol{v}_t = log(\boldsymbol{\lambda}_t)$.

Follow the notation in Liu (2012), let $\boldsymbol{u} = (u_1, u_2) \in [0,1]^2$ be the independent uniform distribution. We can describe the random function $f_{\boldsymbol{u}}(\boldsymbol{v})$, where $\{\boldsymbol{v}_t, t \geq 1\}$ is a bivariate Markov Chain, as

$$f_{\boldsymbol{u}}(\boldsymbol{v}) = \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v} + \boldsymbol{B}log(F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) + \boldsymbol{e}) \tag{6.2.2}$$

where $F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) = (F_{v_1}^{-1}(u_1), F_{v_2}^{-1}(u_2))^T$ and $F_v^{-1}(u) = \inf\{y \geq 0 : F_v(y) \geq u\}$. From 6.2.2, for all time t, $\{\boldsymbol{u}_t, t \geq 1\}$ is independent uniform distribution on $[0,1]^2$, so we have $\boldsymbol{v}_t = f_{\boldsymbol{u}_t}(\boldsymbol{v}_{t-1})$. Combining this bivariate log-Poisson model with Liu (2012), we have the following propositions:

Proposition 6.2.1. *Assume model 6.1.2, $\boldsymbol{D}$ is a 2×1 vector with nonnegative entries, $\boldsymbol{A}$ and $\boldsymbol{B}$ are both 2×2 matrix with entries $\in \mathbb{R}$ (i.e. both positive and negative entries can be considered), and suppose that $\rho(\boldsymbol{A}) < 1$.*

1. If $\rho(\boldsymbol{A} + \boldsymbol{B}) < 1$ and $\rho(\boldsymbol{A} - \boldsymbol{B}) < 1$, then there exists at least one stationary distribution to $\{\boldsymbol{v}_t\}$. In addition, the stationary is unique if $||\boldsymbol{A}||_p \leq 1$ for some $1 \leq p \leq \infty$.

2. For some $1 \leq p \leq \infty$, if $||\boldsymbol{A}||_p + 2^{(1-1/p)}||\boldsymbol{B}||_p < 1$, then $\{\boldsymbol{v}_t\}$ is a GMC Markov chain with a unique stationary and ergodic distribution, denoted by π.

Proof. 1. From the definition in Meyn and Tweedie (2005) and in Liu (2012), $\{\boldsymbol{v}_t\}$ is a weak Feller chain, i.e., for every sequence $\{x_n\}$ in $\boldsymbol{X}$, such that $x_n \to x \in \boldsymbol{X}$, $Pf(x_n) \to Pf(x)$, whenever $f \in C_b(\boldsymbol{X})$, f is any bounded continuous function on $[0,\infty) \times [0,\infty)$. The proof is similar as Liu (2012). For simplicity, first consider K=2.

For $\boldsymbol{v} = (v_1, v_2)^T$, where $v_1 \equiv log\lambda_1$, $v_2 \equiv log\lambda_2$, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$, let $\boldsymbol{v}_1$ represents $\boldsymbol{v}$ at time 1, we have

$$P_{\boldsymbol{v}} f = \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\boldsymbol{y} + \boldsymbol{e}) \big| \boldsymbol{v}_1 = \boldsymbol{v}) p(\boldsymbol{y}|\boldsymbol{v}) \tag{6.2.3}$$

where $\boldsymbol{y} = (y_1, y_2)^T$, $\boldsymbol{e} = (1, 1)^T$, and $p(\boldsymbol{y}|\boldsymbol{v})$ is the pmf of $(Y_1, Y_2)^T$, with $Y_{t,1} = N(\lambda_{t,1})$, $Y_{t,2} = N(\lambda_{t,2})$ and $Y_{t,1}$, $Y_{t,2}$ are conditional independent. So $P\boldsymbol{v}f$ is continuous by following the continuity of f. Due to *Theorem* 12.0.1 (i) in Meyn and Tweedie (2005), if we can prove $\boldsymbol{v}$ is bounded in probability on average, i.e., for any $\boldsymbol{v}_1 \in \boldsymbol{X}$, and $\epsilon > 0$, there exists $C = [0, M_1] \times [0, M_2] \in \mathbb{R}^2$, such that $\frac{1}{m} \sum_{t=1}^{m} P^t(\boldsymbol{v}_1, C) \geq 1 - \epsilon$, for all $m \geq 1$, where $P^t(\boldsymbol{v}_1, \cdot)$ is the t-th transition probability of $\{\boldsymbol{v}_t\}$, then there exists at least one invariant probability measure. Since we have $\boldsymbol{v}_2 = \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\boldsymbol{Y}_1 + \boldsymbol{e})$, where the subscript number represent time sequence, then

$$\mathbb{E}[\boldsymbol{v}_2|\boldsymbol{v}_1] = \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}\mathbb{E}[log(\boldsymbol{Y}_1 + \boldsymbol{e})|\boldsymbol{v}_1] \tag{6.2.4}$$

by Jensen's Inequality, we have

$$\boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}\mathbb{E}[log(\boldsymbol{Y}_1 + \boldsymbol{e})|\boldsymbol{v}_1] \leq \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\mathbb{E}[\boldsymbol{Y}_1 + \boldsymbol{e}|\boldsymbol{v}_1])$$

then 6.2.4 satisfy

$$\mathbb{E}[\boldsymbol{v}_2|\boldsymbol{v}_1] \quad \leq \quad \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\mathbb{E}[\boldsymbol{Y}_1 + \boldsymbol{e}|\boldsymbol{v}_1])$$

$$
\begin{aligned}
&= \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\mathbb{E}[\boldsymbol{Y}_1|\boldsymbol{v}_1] + \boldsymbol{e}) \\
&= \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\boldsymbol{\lambda}_1 + \boldsymbol{e}) \qquad (6.2.5)
\end{aligned}
$$

where $\boldsymbol{\lambda}_1 \equiv exp(\boldsymbol{v}_1)$.

For $\boldsymbol{\lambda}_1 \in [1, \infty)^2$, i.e., $\boldsymbol{v}_1 \equiv log(\boldsymbol{\lambda}_1) \in [0, \infty)^2$, 6.2.5 satisfy

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{v}_2|\boldsymbol{v}_1] &\leq \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}\boldsymbol{v}_1 + \boldsymbol{B}\boldsymbol{e} \\
&= \boldsymbol{D} + (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{v}_1 + \boldsymbol{B}\boldsymbol{e} \qquad (6.2.6)
\end{aligned}
$$

by induction, for any $t \geq 1$,

$$
\mathbb{E}[\boldsymbol{v}_{t+1}|\boldsymbol{v}_1] \leq [\boldsymbol{I} + (\boldsymbol{A} + \boldsymbol{B}) + \cdots + (\boldsymbol{A} + \boldsymbol{B})^{t-1}](\boldsymbol{D} + \boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A} + \boldsymbol{B})^t\boldsymbol{v}_1 \qquad (6.2.7)
$$

where $\boldsymbol{I}$ is a 2×2 identity matrix, $\boldsymbol{e} = (1, 1)^T$. So when $\rho(\boldsymbol{A} + \boldsymbol{B}) < 1$, we have $(\boldsymbol{A} + \boldsymbol{B})^t \to 0$ as $t \to 0$, also $I - (\boldsymbol{A} + \boldsymbol{B})$ is nonsingular and $\sum_{t=0}^{\infty}(\boldsymbol{A} + \boldsymbol{B})^t = [\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{B})]^{-1}$. Then 6.2.7 becomes

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{v}_{t+1}|\boldsymbol{v}_1] &\leq [\boldsymbol{I} + (\boldsymbol{A} + \boldsymbol{B}) + \cdots + (\boldsymbol{A} + \boldsymbol{B})^{t-1}](\boldsymbol{D} + \boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{v}_1 \\
&= [\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{B})]^{-1}(\boldsymbol{D} + \boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{v}_1 \qquad (6.2.8)
\end{aligned}
$$

This follows that

$$\begin{aligned} P^t(\boldsymbol{v}_{t+1} \in C|\boldsymbol{v}_1) &= P(v_{t+1,1} \leq M_1, v_{t+1,2} \leq M_2|\boldsymbol{v}_1) \\ &\geq 1 - P(v_{t+1,1} > M_1|\boldsymbol{v}_1) - P(v_{t+1,2} > M_2|\boldsymbol{v}_1) \end{aligned} \tag{6.2.9}$$

from Markov Inequality, i.e., $x > 0$ for any $\epsilon > 0$, $P(x > \epsilon) \leq \frac{E[x]}{\epsilon}$, we have $P(v_{t+1,1} > M_1|\boldsymbol{v}_1) \leq \frac{E[v_{t+1,1}|\boldsymbol{v}_1]}{M_1}$, where $v_{t+1,1} > 0$, $M_1 > 0$, then 6.2.9 satisfies

$$\begin{aligned} P^t(\boldsymbol{v}_{t+1} \in C|\boldsymbol{v}_1) &\geq 1 - \frac{\mathbb{E}[v_{t+1,1}|\boldsymbol{v}_1]}{M_1} - \frac{\mathbb{E}[v_{t+1,2}|\boldsymbol{v}_1]}{M_2} \\ &= 1 - \boldsymbol{M}^T\mathbb{E}[\boldsymbol{v}_{t+1}|\boldsymbol{v}_1] \\ &\geq 1 - \boldsymbol{M}^T\{[\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{B})]^{-1}(\boldsymbol{D} + \boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{v}_1\} \end{aligned} \tag{6.2.10}$$

where $\boldsymbol{M}^T = (\frac{1}{M_1} \quad \frac{1}{M_2})$. There exists M_1, $M_2 \in \mathbb{R}$ large enough to such that $\boldsymbol{M}^T\{[\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{B})]^{-1}(\boldsymbol{D} + \boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{v}_1\} \leq \epsilon$, which means that $P^t(\boldsymbol{v}_{t+1} \in C|\boldsymbol{v}_1) \geq 1 - \epsilon$ for all $t \geq 1$.

For $\boldsymbol{\lambda}_1 \in (0,1)^2$, i.e., $\boldsymbol{v}_1 \equiv log(\boldsymbol{\lambda}_1) \in (-\infty, 0)^2$, 6.2.5 satisfy

$$\begin{aligned} \mathbb{E}[\boldsymbol{v}_2|\boldsymbol{v}_1] &= \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}log(\boldsymbol{\lambda}_1 + \boldsymbol{e}) \\ &\leq \boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_1 + \boldsymbol{B}(-log\boldsymbol{\lambda}_1 + \boldsymbol{e}) \\ &= \boldsymbol{D} + (\boldsymbol{A} - \boldsymbol{B})\boldsymbol{v}_1 + \boldsymbol{B}\boldsymbol{e} \end{aligned} \tag{6.2.11}$$

by induction for any $t \geq 1$,

$$\mathbb{E}[\boldsymbol{v}_{t+1}|\boldsymbol{v}_1] \leq [\boldsymbol{I} + (\boldsymbol{A}-\boldsymbol{B}) + \cdots + (\boldsymbol{A}-\boldsymbol{B})^{t-1}](\boldsymbol{D}+\boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A}-\boldsymbol{B})^t\boldsymbol{v}_1 \quad (6.2.12)$$

So when $\rho(\boldsymbol{A}-\boldsymbol{B}) < 1$, we have $(\boldsymbol{A}-\boldsymbol{B})^t \to 0$ as $t \to 0$, also $I-(\boldsymbol{A}+\boldsymbol{B})$ is nonsingular and $\sum_{t=0}^{\infty}(\boldsymbol{A}-\boldsymbol{B})^t = [\boldsymbol{I}-(\boldsymbol{A}-\boldsymbol{B})]^{-1}$. Then 6.2.12 becomes

$$\begin{aligned}
\mathbb{E}[\boldsymbol{v}_{t+1}|\boldsymbol{v}_1] &\leq [\boldsymbol{I} + (\boldsymbol{A}-\boldsymbol{B}) + \cdots + (\boldsymbol{A}-\boldsymbol{B})^{t-1}](\boldsymbol{D}+\boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}_1 \\
&= [\boldsymbol{I}-(\boldsymbol{A}-\boldsymbol{B})]^{-1}(\boldsymbol{D}+\boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}_1 \qquad (6.2.13)
\end{aligned}$$

Same as $\boldsymbol{\lambda}_1 \in [1,\infty)^2$, we have

$$\begin{aligned}
P^t(\boldsymbol{v}_{t+1} \in C|\boldsymbol{v}_1) &= P(v_{t+1,1} \leq M_1, v_{t+1,2} \leq M_2|\boldsymbol{v}_1) \\
&\geq 1 - P(v_{t+1,1} > M_1|\boldsymbol{v}_1) - P(v_{t+1,2} > M_2|\boldsymbol{v}_1) \\
&\geq 1 - \frac{\mathbb{E}[v_{t+1,1}|\boldsymbol{v}_1]}{M_1} - \frac{\mathbb{E}[v_{t+1,2}|\boldsymbol{v}_1]}{M_2} \\
&= 1 - \boldsymbol{M}^T\mathbb{E}[\boldsymbol{v}_{t+1}|\boldsymbol{v}_1] \\
&\geq 1 - \boldsymbol{M}^T\{[\boldsymbol{I}-(\boldsymbol{A}-\boldsymbol{B})]^{-1}(\boldsymbol{D}+\boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}_1\}
\end{aligned}$$

and there exists M_1, $M_2 \in \mathbb{R}$ large enough to such that $\boldsymbol{M}^T\{[\boldsymbol{I}-(\boldsymbol{A}-\boldsymbol{B})]^{-1}(\boldsymbol{D}+\boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}_1\} \leq \epsilon$, which means that $P^t(\boldsymbol{v}_{t+1} \in C|\boldsymbol{v}_1) \geq 1-\epsilon$ for all $t \geq 1$.

So for all $m \geq 1$, we have $\frac{1}{m}\sum_{t=1}^{m} P^t(\boldsymbol{v}_1, C) \geq 1-\epsilon$. This proves that $\{\boldsymbol{v}_t\}$ is bounded in probability on average. Therefore, $\{\boldsymbol{v}_t\}$ has at least one stationary distribution.

Assume $||\boldsymbol{A}||_p < 1$ for some $1 \leq p \leq \infty$, from equation 6.1.3, we have

$$\boldsymbol{v}_t = (\boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^{t-1})\boldsymbol{D} + \boldsymbol{A}^t\boldsymbol{v}_0 + \sum_{i=0}^{t-1} \boldsymbol{A}^i\boldsymbol{B}log(\boldsymbol{Y}_{t-i-1} + \boldsymbol{e}) \tag{6.2.14}$$

where $\boldsymbol{e} = (1,1)^T$. If $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1} = \cdots = \boldsymbol{0}$ for $t \to \infty$, $t \in \mathbb{N}$, we can see that $(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{D}$ is a reachable state. Due to Theorem 18.4.4 in Meyn and Tweedie (2005), if $\{\boldsymbol{v}_t\}$ is an e-chain, i.e., for each continuous function f on compact sets $[0,\infty) \times [0,\infty)$, and $\epsilon > 0$, let $||\cdot||$ is some norm on $\mathbb{R}^2$, there exists an $\eta > 0$, s.t. $|P^n_{\boldsymbol{x}_1} f - P^n_{\boldsymbol{z}_1} f| < \epsilon$, for $||\boldsymbol{x}_1 - \boldsymbol{z}_1|| < \eta$ for all $n \geq 1$, where $\boldsymbol{x}_1 = (x_{1,1}, x_{1,2})^T$, $\boldsymbol{z}_1 = (z_{1,1}, z_{1,2})^T$, then the stationary distribution to $\{\boldsymbol{v}_t\}$ is unique.

Based on Liu (2012), assume $|f| < 1$, let ϵ' and η small enough to make sure $\epsilon' + 8\eta/(1 - ||\boldsymbol{A}||_p) < \epsilon$, $|f(\boldsymbol{x}_1) - f(\boldsymbol{z}_1)| < \epsilon'$, whenever $||\boldsymbol{x}_1 - \boldsymbol{z}_1||_p < \eta$ for some $1 \leq p \leq \infty$.

For $n = 1$, from 6.2.3, we have

$$\begin{aligned}
|P_{\boldsymbol{x}_1} f - P_{\boldsymbol{z}_1} f| = \Big| \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \big[& f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}log(\boldsymbol{y} + \boldsymbol{e}))p(\boldsymbol{y}|\boldsymbol{x}_1) \\
& - f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}log(\boldsymbol{y} + \boldsymbol{e}))p(\boldsymbol{y}|\boldsymbol{z}_1) \big] \Big| \\
\leq & \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} p(\boldsymbol{y}|\boldsymbol{x}_1) \big| f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}log(\boldsymbol{y} + \boldsymbol{e})) - f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}log(\boldsymbol{y} + \boldsymbol{e})) \big| \\
& + \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \big| p(\boldsymbol{y}|\boldsymbol{x}_1) - p(\boldsymbol{y}|\boldsymbol{z}_1) \big| \times \big| f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}log(\boldsymbol{y} + \boldsymbol{e})) \big|
\end{aligned}$$

$$= I + II \tag{6.2.15}$$

where $\boldsymbol{y} = (y_1, y_2)^T$, and $\boldsymbol{e} = (1,1)^T$. First focus on II in 6.2.15, since $|f| < 1$, we have

$$\begin{aligned}
II \leq & \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} \big|p(\boldsymbol{y}|\boldsymbol{x}_1) - p(\boldsymbol{y}|\boldsymbol{z}_1)\big| \\
= & \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} \big|P(y_1|x_{1,1})P(y_2|x_{1,2}) - P(y_1|z_{1,1})P(y_2|z_{1,2})\big| \\
\leq & \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} \big|P(y_1|x_{1,1}) - P(y_1|z_{1,1})\big|P(y_2|x_{1,2}) \\
& + \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} P(y_1|z_{1,1})\big|P(y_2|x_{1,2}) - P(y_2|z_{1,2})\big| \\
\leq & \sum_{i=0}^{\infty} \big|P(i|x_{1,1}) - P(i|z_{1,1})\big| + \sum_{i=0}^{\infty} \big|P(i|x_{1,2}) + P(i|z_{1,2})\big|
\end{aligned} \tag{6.2.16}$$

since $y_1 \sim \text{Poisson}(\lambda_1)$, $y_2 \sim \text{Poisson}(\lambda_2)$, we have

$$\begin{aligned}
\sum_{i=0}^{\infty} \big|P(i|x_{1,1}) - P(i|z_{1,1})\big| = & \sum_{i=0}^{\infty} \Big|\frac{e^{-x_{1,1}}x_{1,1}^i}{i!} - \frac{e^{-z_{1,1}}z_{1,1}^i}{i!}\Big| \\
\leq & \sum_{i=0}^{\infty} \big|e^{-x_{1,1}} - e^{-z_{1,1}}\big|\frac{x_{1,1}^i}{i!} + e^{-z_{1,1}}\sum_{i=0}^{\infty}\frac{\big|x_{1,1}^i - z_{1,1}^i\big|}{i!} \\
\leq & \, 2(1 - e^{-|x_{1,1}-z_{1,1}|})
\end{aligned} \tag{6.2.17}$$

then combine with equation 6.2.16, we have

$$\sum_{i=0}^{\infty}\left|P(i|x_{1,1})-P(i|z_{1,1})\right|+\sum_{i=0}^{\infty}\left|P(i|x_{1,2})+P(i|z_{1,2})\right|\leq\ 2(1-e^{-|x_{1,1}-z_{1,1}|})+2(1-e^{-|x_{1,2}-z_{1,2}|}) \tag{6.2.18}$$

From the definition of vector norm and p-induced norm, we have $|x_{1,i}-z_{1,i}|\leq||\boldsymbol{x}_1-\boldsymbol{z}_1||_1=|x_{1,1}-z_{1,1}|+|x_{1,2}-z_{1,2}|\leq C_p||\boldsymbol{x}_1-\boldsymbol{z}_1||_p$ for $i=1,2,\ 1\leq p\leq\infty$, where $C_p=2^{1-\frac{1}{p}}\leq 2$ from the inequality of vector norm. Then 6.2.16 becomes

$$\begin{aligned}\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\left|p(\boldsymbol{y}|\boldsymbol{x}_1)-p(\boldsymbol{y}|\boldsymbol{z}_1)\right|&\leq 2(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})+2(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})\\&=4(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})\end{aligned} \tag{6.2.19}$$

from 6.2.16, we have proved $II\leq\ 4(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})$.

For I, since $||\boldsymbol{x}_1-\boldsymbol{z}_1||_p\leq\eta$, and $||\boldsymbol{A}||_p\leq 1$, we have

$$\begin{aligned}&||(\boldsymbol{D}+\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{B}log(\boldsymbol{y}+\boldsymbol{e}))-(\boldsymbol{D}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}log(\boldsymbol{y}+\boldsymbol{e}))||_p\\&=||\boldsymbol{A}(\boldsymbol{x}_1-\boldsymbol{z}_1)||_p\leq||\boldsymbol{A}||_p||\boldsymbol{x}_1-\boldsymbol{z}_1||_p\leq\eta\end{aligned} \tag{6.2.20}$$

then from the assumption stated previously, we have $I=\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}p(\boldsymbol{y}|\boldsymbol{x}_1)\left|f(\boldsymbol{D}+\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{B}log(\boldsymbol{y}+\boldsymbol{e}))-f(\boldsymbol{D}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}log(\boldsymbol{y}+\boldsymbol{e}))\right|\leq|f(\boldsymbol{x}_1-\boldsymbol{z}_1)|\leq\epsilon'$.

Combining the results for I and II, we have

$$|P_{\boldsymbol{x}_1} f - P_{\boldsymbol{z}_1} f| \leq I + II \leq \epsilon' + 4(1 - e^{-2||\boldsymbol{x}_1 - \boldsymbol{z}_1||_p}) \tag{6.2.21}$$

For the case $n = 2$, we have

$$\begin{aligned} |P^2_{\boldsymbol{x}_1} f - P^2_{\boldsymbol{z}_1} f| &= \Big| \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \big[P(y_1, y_2|\boldsymbol{x}_1) P_{\boldsymbol{x}_2} f - P(y_1, y_2|\boldsymbol{z}_1) P_{\boldsymbol{z}_2} f \big] \Big| \\ &\leq \sum_{y_1=0}^{\infty} \sum_{y_1=0}^{\infty} P(y_1, y_2|\boldsymbol{x}_1) \big| P_{\boldsymbol{x}_2} f - P_{\boldsymbol{z}_2} f \big| + \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \big| P(y_1, y_2|\boldsymbol{x}_1) - P(y_1, y_2|\boldsymbol{z}_1) \big| \times \big| P_{\boldsymbol{z}_2} f \big| \end{aligned} \tag{6.2.22}$$

in which

$$\begin{aligned} \big| P_{\boldsymbol{x}_2} f - P_{\boldsymbol{z}_2} f \big| \leq & \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} P(y_1, y_2|\boldsymbol{x}_2) \big| f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{B} log(\boldsymbol{y} + \boldsymbol{e})) - f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{z}_2 + \boldsymbol{B} log(\boldsymbol{y} + \boldsymbol{e})) \big| \\ & + \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \big| P(y_1, y_2|\boldsymbol{x}_2) - P(y_1, y_2|\boldsymbol{z}_2) \big| \times \big| f(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{z}_2 + \boldsymbol{B} log(\boldsymbol{y} + \boldsymbol{e})) \big| \end{aligned} \tag{6.2.23}$$

since $\boldsymbol{x}_2 = \boldsymbol{D} + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B} log(\boldsymbol{y} + \boldsymbol{e})$, $\boldsymbol{z}_2 = \boldsymbol{D} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B} log(\boldsymbol{y} + \boldsymbol{e})$, then $||\boldsymbol{x}_2 - \boldsymbol{z}_2||_p = ||\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{z}_1)||_p \leq ||\boldsymbol{A}||_p ||\boldsymbol{x}_1 - \boldsymbol{z}_1||_p \leq \eta$, then $|f(\boldsymbol{x}_2 - f(\boldsymbol{z}_2)| < \epsilon'$, and follow 6.2.21, we have

$$\big| P_{\boldsymbol{x}_2} f - P_{\boldsymbol{z}_2} f \big| \leq I + II \leq \epsilon' + 4(1 - e^{-2||\boldsymbol{x}_2 - \boldsymbol{z}_2||_p}) \tag{6.2.24}$$

and from 6.2.19, we have

$$\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\left|P(y_1,y_2|\boldsymbol{x}_1)-P(y_1,y_2|\boldsymbol{z}_1)\right|\times\left|P_{\boldsymbol{z}_2}f\right|\leq\ 4(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p}) \tag{6.2.25}$$

substitute 6.2.24, 6.2.25 into 6.2.22, we have

$$\begin{aligned}|P^2_{\boldsymbol{x}_1}f-P^2_{\boldsymbol{z}_1}f|&\leq\epsilon'+4(1-e^{-2||\boldsymbol{x}_2-\boldsymbol{z}_2||_p})+4(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})\\&\leq\epsilon'+4(1-e^{-2||\boldsymbol{A}||_p||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})+4(1-e^{-2||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})\end{aligned} \tag{6.2.26}$$

By induction and the inequality $e^x\geq x+1$, since $||\boldsymbol{x}_1-\boldsymbol{z}_1||_p<\eta$, so for $n\geq 1$, we have

$$\begin{aligned}|P^n_{\boldsymbol{x}_1}f-P^n_{\boldsymbol{z}_1}f|&\leq\epsilon'+4\sum_{i=0}^{n-1}(1-e^{-2||\boldsymbol{A}||_p^i||\boldsymbol{x}_1-\boldsymbol{z}_1||_p})\\&\leq\epsilon'+4\sum_{i=0}^{\infty}(2||\boldsymbol{A}||_p^i||\boldsymbol{x}_1-\boldsymbol{z}_1||_p)\\&=\epsilon'+8\sum_{i=0}^{\infty}||\boldsymbol{A}||_p^i||\boldsymbol{x}_1-\boldsymbol{z}_1||_p\\&\leq\epsilon'+\frac{8\eta}{1-||\boldsymbol{A}||_p}\\&<\epsilon\end{aligned} \tag{6.2.27}$$

which proves that $\{\boldsymbol{v}_t\}$ is an e-chain. Combining with $\{\boldsymbol{v}_t\}$ is bounded in probability on average, and a reachable state $(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{D}$ exists, according to the Theorem 18.4.4 in Meyn and Tweedie (2005), there exists a unique stationary distribution to $\{\boldsymbol{v}_t\}$.

2. To prove proposition 2, we need to follow the Theorem 2 in Wu and Shao (2004). Condition 1 in Wu and Shao (2004) provides a bound on the intercept of the random transform F, so it holds trivially. For condition 2 in Wu and Shao (2004), similar as Liu (2012), consider $\{\boldsymbol{v}_0\} = (v_{0,1}, v_{0,2})^T$ fixed at time 0, for any time t, $\boldsymbol{v} = (v_1, v_2)^T \in \boldsymbol{X}$, where $\boldsymbol{X}$ is the state space, then

$$\begin{aligned}
&\mathbb{E}||\boldsymbol{v}_1(\boldsymbol{v}) - \boldsymbol{v}_1(\boldsymbol{v}_0)||_p \\
&= \int ||(\boldsymbol{D} + \boldsymbol{A}\boldsymbol{v} + \boldsymbol{B}log(F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) + \boldsymbol{e})) - (\boldsymbol{D} + \boldsymbol{A}\boldsymbol{v}_0 + \boldsymbol{B}log(F_{\boldsymbol{v}_0}^{-1}(\boldsymbol{u}) + \boldsymbol{e}))||_p d(\boldsymbol{u}) \\
&\leq ||\boldsymbol{A}(\boldsymbol{v} - \boldsymbol{v}_0)||_p + \int ||\boldsymbol{B}[log(F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) + \boldsymbol{e}) - log(F_{\boldsymbol{v}_0}^{-1}(\boldsymbol{u}) + \boldsymbol{e})]||_p d(\boldsymbol{u}) \\
&\leq ||\boldsymbol{A}||_p ||\boldsymbol{v} - \boldsymbol{v}_0||_p + ||\boldsymbol{B}||_p \int ||log(F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) + \boldsymbol{e}) - log(F_{\boldsymbol{v}_0}^{-1}(\boldsymbol{u}) + \boldsymbol{e})||_1 d(\boldsymbol{u}) \quad (6.2.28)
\end{aligned}$$

since

$$\begin{aligned}
&\int ||log(F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) + \boldsymbol{e}) - log(F_{\boldsymbol{v}_0}^{-1}(\boldsymbol{u}) + \boldsymbol{e})||_1 d(\boldsymbol{u}) \\
&= \int \left[|log(F_{v_1}^{-1}(u_1) + 1) - log(F_{v_{0,1}}^{-1}(u_1) + 1)| + |log(F_{v_2}^{-1}(u_2) + 1) - log(F_{v_{0,2}}^{-1}(u_2) + 1)|\right] d(\boldsymbol{u}) \\
&= \mathbb{E}|log(F_{v_1}^{-1}(u_1) + 1) - log(F_{v_{0,1}}^{-1}(u_1) + 1)| + \mathbb{E}|log(F_{v_2}^{-1}(u_2) + 1) - log(F_{v_{0,2}}^{-1}(u_2) + 1)| \\
&\qquad\qquad (6.2.29)
\end{aligned}$$

from Lamma A.1 in Fokianos and Tjøstheim (2011), 6.2.29 $\rightarrow$ $|v_1 - v_{0,1}| + |v_2 - v_{0,2}| = ||\boldsymbol{v} - \boldsymbol{v}_0||_1$, as $v \rightarrow \infty$, so we have

$$\int ||log(F_{\boldsymbol{v}}^{-1}(\boldsymbol{u}) + \boldsymbol{e}) - log(F_{\boldsymbol{v}_0}^{-1}(\boldsymbol{u}) + \boldsymbol{e})||_1 d(\boldsymbol{u}) = ||\boldsymbol{v} - \boldsymbol{v}_0||_1 \qquad (6.2.30)$$

Again from the inequality of vector norm, $||\boldsymbol{v} - \boldsymbol{v}_0||_1 \leq 2^{1-\frac{1}{p}}||\boldsymbol{v} - \boldsymbol{v}_0||_p$, then

$$\begin{aligned}
\mathbb{E}||f_{\boldsymbol{u}}(\boldsymbol{v}) - f_{\boldsymbol{u}}(\boldsymbol{v}_0)||_p &\leq ||\boldsymbol{A}||_p||\boldsymbol{v} - \boldsymbol{v}_0||_p + ||\boldsymbol{B}||_p||\boldsymbol{v} - \boldsymbol{v}_0||_1 \\
&\leq ||\boldsymbol{A}||_p||\boldsymbol{v} - \boldsymbol{v}_0||_p + 2^{1-\frac{1}{p}}||\boldsymbol{B}||_p||\boldsymbol{v} - \boldsymbol{v}_0||_p \\
&\leq (||\boldsymbol{A}||_p + 2^{1-\frac{1}{p}}||\boldsymbol{B}||_p)||\boldsymbol{v} - \boldsymbol{v}_0||_p \qquad (6.2.31)
\end{aligned}$$

Due to the definition and Condition 2 in Wu and Shao (2004), we have proved that $\{\boldsymbol{v}_t\}$ is geometric contracting under $(||\boldsymbol{A}||_p + 2^{1-\frac{1}{p}}||\boldsymbol{B}||_p) < 1$, and hence $\{\boldsymbol{v}_t\}$ has a unique stationary distribution. Based on Liu (2012) and according to Theorem 3.1, Remark 3.1 and Corollary 3.1 in Doukhan and Wintenberger (2008), we know that the stationary distribution is also τ-weakly dependent and automatically an ergodic process as a causal Bernoulli shift solution.

□

Chapter 7

Likelihood Inference

7.1 Likelihood Function and Score Function

For simplicity, consider the bivariate log-linear Poisson autoregression model, we denote $\boldsymbol{\theta}$ as a ten-dimensional vector of unknown parameters, i.e.
$\boldsymbol{\theta} = (d_1, d_2, a_{11}, a_{12}, a_{21}, a_{22}, b_{11}, b_{12}, b_{21}, b_{22})$. Then the conditional likelihood function for $\boldsymbol{\theta}$ given the observations $\{Y_{t,1}, Y_{t-1,1}, ..., Y_{0,1}\}$ and $\{Y_{t,2}, Y_{t-1,2}, ..., Y_{0,2}\}$ with the starting value $\lambda_{0,1} = exp(v_{0,1})$ and $\lambda_{0,2} = exp(v_{0,2})$ is:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n} \left[\frac{exp(-\lambda_{t,1}(\boldsymbol{\theta}))\lambda_{t,1}(\boldsymbol{\theta})^{Y_{t,1}}}{Y_{t,1}!} \times \frac{exp(-\lambda_{t,2}(\boldsymbol{\theta}))\lambda_{t,2}(\boldsymbol{\theta})^{Y_{t,2}}}{Y_{t,2}!} \right] \tag{7.1.1}$$

Here we use the Poisson assumption

$$\begin{pmatrix} \lambda_{t,1}(\boldsymbol{\theta}) \\ \lambda_{t,2}(\boldsymbol{\theta}) \end{pmatrix} = \boldsymbol{D} + \boldsymbol{A} \begin{pmatrix} \lambda_{t-1,1}(\boldsymbol{\theta}) \\ \lambda_{t-1,2}(\boldsymbol{\theta}) \end{pmatrix} + \boldsymbol{B} \begin{pmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{pmatrix} \tag{7.1.2}$$

by 6.1.3, and the log-likelihood function is given by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left[Y_{t,1} v_{t,1}(\boldsymbol{\theta}) - exp(v_{t,1}(\boldsymbol{\theta})) + Y_{t,2} v_{t,2}(\boldsymbol{\theta}) - exp(v_{t,2}(\boldsymbol{\theta})) \right] \tag{7.1.3}$$

where

$$v_{t,1} = d_1 + a_{11} v_{t-1,1} + a_{12} v_{t-1,2} + b_{11}(log(Y_{t-1,1} + 1)) + b_{12}(log(Y_{t-1,2} + 1)) \tag{7.1.4}$$

$$v_{t,2} = d_2 + a_{21} v_{t-1,1} + a_{22} v_{t-1,2} + b_{21}(log(Y_{t-1,1} + 1)) + b_{22}(log(Y_{t-1,2} + 1)) \tag{7.1.5}$$

for the simplicity and expansion formula for model 6.1.3. And the score function is:

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \left[Y_{t,1} - exp(v_{t,1}(\boldsymbol{\theta})) \frac{\partial v_{t,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + Y_{t,2} - exp(v_{t,2}(\boldsymbol{\theta})) \frac{\partial v_{t,2}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \tag{7.1.6}$$

where the partial derivative of $\boldsymbol{v}_t(\boldsymbol{\theta})$ to $\boldsymbol{\theta}$ are ten-dimensional vectors with notation $\boldsymbol{v}_t(\boldsymbol{\theta}) = (v_{t,1}(\boldsymbol{\theta}), v_{t,2}(\boldsymbol{\theta}))^T$ and $\boldsymbol{v}_{t-1}(\boldsymbol{\theta}) = (v_{t-1,1}(\boldsymbol{\theta}), v_{t-1,2}(\boldsymbol{\theta}))^T$ are given by

$$\frac{\partial \boldsymbol{v}_t}{\partial d_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \boldsymbol{A} \frac{\partial \boldsymbol{v}_{t-1}}{\partial d_1}, \qquad \frac{\partial \boldsymbol{v}_t}{\partial d_2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \boldsymbol{A} \frac{\partial \boldsymbol{v}_{t-1}}{\partial d_2}$$

$$\frac{\partial \boldsymbol{v}_t}{\partial a_{11}} = \begin{pmatrix} v_{t-1,1} \\ 0 \end{pmatrix} + \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial a_{11}}, \qquad \frac{\partial \boldsymbol{v}_t}{\partial a_{12}} = \begin{pmatrix} v_{t-1,2} \\ 0 \end{pmatrix} + \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial a_{12}}$$

$$\frac{\partial \boldsymbol{v}_t}{\partial a_{21}} = \begin{pmatrix} 0 \\ v_{t-1,1} \end{pmatrix} + \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial a_{21}}, \qquad \frac{\partial \boldsymbol{v}_t}{\partial a_{22}} = \begin{pmatrix} 0 \\ v_{t-1,2} \end{pmatrix} + \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial a_{22}}$$

$$\frac{\partial \boldsymbol{v}_t}{\partial b_{11}} = \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial b_{11}} + \begin{pmatrix} log(Y_{t-1,1}+1) \\ 0 \end{pmatrix}, \qquad \frac{\partial \boldsymbol{v}_t}{\partial a_{12}} = \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial b_{12}} + \begin{pmatrix} log(Y_{t-1,2}+1) \\ 0 \end{pmatrix}$$

$$\frac{\partial \boldsymbol{v}_t}{\partial b_{21}} = \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial b_{21}} + \begin{pmatrix} 0 \\ log(Y_{t-1,1}+1) \end{pmatrix}, \qquad \frac{\partial \boldsymbol{v}_t}{\partial a_{22}} = \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial b_{22}} + \begin{pmatrix} 0 \\ log(Y_{t-1,2}+1) \end{pmatrix} \tag{7.1.7}$$

Here, for simplicity, we rewrite the derivative $\partial \boldsymbol{v}_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ with k-dimensional observations. Define vector $\boldsymbol{e}_i = (0, 0, \ldots, 1, 0, \ldots, 0)^T$ with the i-th element equals to 1 while the others are 0 for the total of k elements in the vector. Then the derivative of $\boldsymbol{v}_t(\boldsymbol{\theta})$ to the element in the vector of $\boldsymbol{D}$ is:

$$\frac{\partial \boldsymbol{v}_t}{\partial d_i} = \boldsymbol{e}_i + \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial d_i} \tag{7.1.8}$$

for $i = 1, 2, \ldots, k$.

For i in $(1, \ldots, k)$, and j in $(1, \ldots, k)$, the derivative of $\boldsymbol{v}_t(\boldsymbol{\theta})$ to the element in the vector of $\boldsymbol{A}$ and to the element in the vector of $\boldsymbol{B}$ are respectively as

$$\frac{\partial \boldsymbol{v}_t}{\partial a_{ij}} = \boldsymbol{v}_{t-1} \cdot \boldsymbol{e}_i + \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial a_{ij}} \tag{7.1.9}$$

and

$$\frac{\partial \boldsymbol{v}_t}{\partial b_{ij}} = \boldsymbol{A}\frac{\partial \boldsymbol{v}_{t-1}}{\partial b_{ij}} + log(Y_{t-1,j} + 1) \cdot \boldsymbol{e}_i \tag{7.1.10}$$

where "$\cdot$" is dot production in function 7.1.9 and function 7.1.10.

We can get the estimation of $\boldsymbol{\theta}$ denoted as $\hat{\boldsymbol{\theta}}$ by the conditional maximum likelihood estimation from the solution of $\boldsymbol{S}_n(\boldsymbol{\theta}) = 0$. Also, the Hessian matrix for model 6.1.3 is

$$\begin{aligned}
\boldsymbol{H}_n(\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \\
&= \sum_{t=1}^{n} exp(v_{t,1}(\boldsymbol{\theta}))\frac{\partial v_{t,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial v_{t,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}' - \sum_{t=1}^{n}(Y_{t,1} - exp(v_{t,1}(\boldsymbol{\theta})))\frac{\partial^2 v_{t,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \\
&+ \sum_{t=1}^{n} exp(v_{t,2}(\boldsymbol{\theta}))\frac{\partial v_{t,2}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial v_{t,2}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}' - \sum_{t=1}^{n}(Y_{t,2} - exp(v_{t,2}(\boldsymbol{\theta})))\frac{\partial^2 v_{t,2}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}
\end{aligned} \tag{7.1.11}$$

where $\boldsymbol{\theta}$ is a ten-dimensional vector.

Chapter 8

Simulation Study

8.1 Simulations for Bivariate Log-linear Poisson Autoregression

We use simulations to illustrate the theory of the bivariate model presented in Chapter 6. The parameters in model 6.1.3 are estimated by direct optimization of the log-likelihood function 7.1.3. Following the method for starting value used in Fokianos and Tjøstheim (2011), we use a routine GLM fit of the model $\begin{pmatrix} \boldsymbol{v}_{t,1} \\ \boldsymbol{v}_{t,2} \end{pmatrix} =$

$\boldsymbol{D}+\boldsymbol{B}\begin{pmatrix} log(\boldsymbol{Y}_{t-1,1}+1) \\ log(\boldsymbol{Y}_{t-1,2}+1) \end{pmatrix}$ for bivariate model by using the recursions function in 7.1.7 to get the initial values of the entries for the parameter vector $\boldsymbol{D}$ and the matrix $\boldsymbol{B}$, and the iterative least squares estimation are used to fit the model (see Fokianos et al.(2009)). With the estimates of $\boldsymbol{D}$ and $\boldsymbol{B}$ which are denoted as $(\widetilde{\boldsymbol{D}}, \widetilde{\boldsymbol{B}})$, and a initial diagonal matrix for $\boldsymbol{A}$ under the constraint, the optimization procedure is taken with the initial values of $(\widetilde{\boldsymbol{D}}, \widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}})$ (see Fokianos and Tjøstheim (2011)).

Following the constraint for the stability and the conditions for bivariate log-liner Poisson autoregression model, we first let

$$\begin{aligned} \boldsymbol{\theta} = (\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{B}) &= \left(\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \right) \\ &= \left(\begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix}, \quad \begin{pmatrix} -0.5 & 0.2 \\ 0.3 & -0.6 \end{pmatrix}, \quad \begin{pmatrix} -0.35 & -0.2 \\ 0.1 & 0.3 \end{pmatrix} \right) \end{aligned}$$

Table 8.1 shows the simulation results for sample size (data points) of 200, 500 ,1000 respectively, and the summary statistics of the sampling distribution of the standardized MLE. All the results are based on 1000 simulations.

Table 8.1: Simulation Results for Bivariate Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (-0.85, -0.3), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.99, 0.06)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
200	d_1	0.5	0.493	0.013	-0.535	4.614	2.76E-06
	d_2	0.8	0.777	0.010	-1.336	6.710	9.08E-05
	a_{11}	-0.5	-0.508	0.012	0.675	4.395	0.002
	a_{12}	0.2	0.174	0.011	0.719	4.479	0.005
	a_{21}	0.3	0.329	0.012	1.666	10.210	4.22E-04
	a_{22}	-0.6	-0.612	0.011	1.788	8.129	2.64E-05
	b_{11}	-0.35	-0.330	0.004	-0.049	3.134	0.914
	b_{12}	-0.2	-0.206	0.004	0.172	3.374	0.219
	b_{21}	0.1	0.101	0.003	-0.067	3.125	0.828
	b_{22}	0.3	0.316	0.003	0.215	3.243	0.341
500	d_1	0.5	0.516	0.009	-0.987	5.765	1.07E-08
	d_2	0.8	0.785	0.008	-1.291	6.749	0.002
	a_{11}	-0.5	-0.517	0.009	0.557	4.900	0.370
	a_{12}	0.2	0.176	0.009	1.083	5.959	7.43E-05
	a_{21}	0.3	0.313	0.008	1.099	6.077	0.008
	a_{22}	-0.6	-0.589	0.008	1.657	8.202	2.91E-04
	b_{11}	-0.35	-0.345	0.003	0.057	2.863	0.314
	b_{12}	-0.2	-0.204	0.002	0.368	3.410	0.432
	b_{21}	0.1	0.103	0.002	-0.022	2.795	0.500
	b_{22}	0.3	0.300	0.002	0.057	3.109	0.432
1000	d_1	0.5	0.490	0.006	-0.506	8.655	4.54E-06
	d_2	0.8	0.786	0.006	-1.407	8.848	0.001
	a_{11}	-0.5	-0.491	0.007	0.066	8.016	0.004
	a_{12}	0.2	0.205	0.007	0.700	9.018	0.010
	a_{21}	0.3	0.313	0.006	1.111	7.760	0.011
	a_{22}	-0.6	-0.588	0.006	1.646	9.677	2.64E-05
	b_{11}	-0.35	-0.348	0.002	-0.016	3.102	0.341
	b_{12}	-0.2	-0.202	0.002	-0.017	3.266	0.610
	b_{21}	0.1	0.101	0.001	-0.035	2.812	0.164
	b_{22}	0.3	0.301	0.001	0.024	2.997	0.914

In table 8.1, the fourth column is the mean of the estimators calculated by the maximum likelihood method. The sample standard deviation of the estimators by the simulation is shown in the fifth column. The sixth column in this table is the sample skewness and the seventh column reports the sample kurtosis. The last column in the table is the p-value of Kolmogorov-Smirnov test statistic for the standardized MLE (testing with the standard normal distribution), as the same statistic considered in Fokianos and Tjøstheim (2011).

In this simulation study, we also plot the histograms and qq-plots for the standardized sample distribution of each entries in $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{D}}, \widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}})$, for which the true values are $\left(\begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix}, \begin{pmatrix} -0.5 & 0.2 \\ 0.3 & -0.6 \end{pmatrix}, \begin{pmatrix} -0.35 & -0.2 \\ 0.1 & 0.3 \end{pmatrix} \right)$. All the plots are for the sample size of 500.

We also change the parameters in $\boldsymbol{D}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ under the constraint $\rho(\boldsymbol{A}) < 1$, $\rho(\boldsymbol{A} + \boldsymbol{B}) < 1$ and $\rho(\boldsymbol{A} - \boldsymbol{B}) < 1$ for more simulations, and the results are in table 8.2, table 8.3, table 8.4, table 8.5.

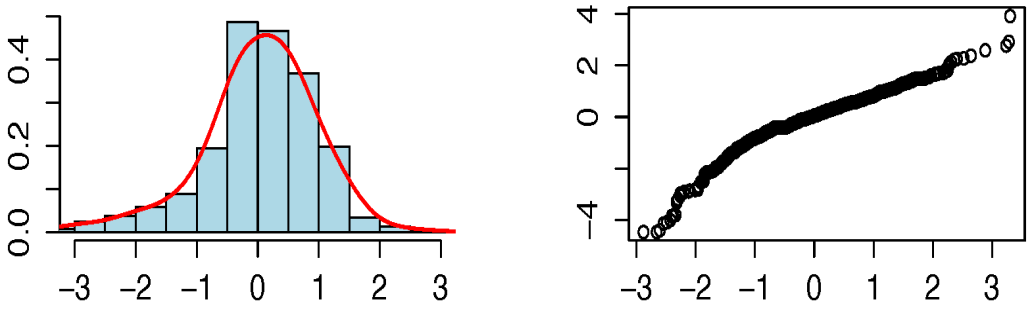

Figure 8.1: Histogram and qq-plots of d_1, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

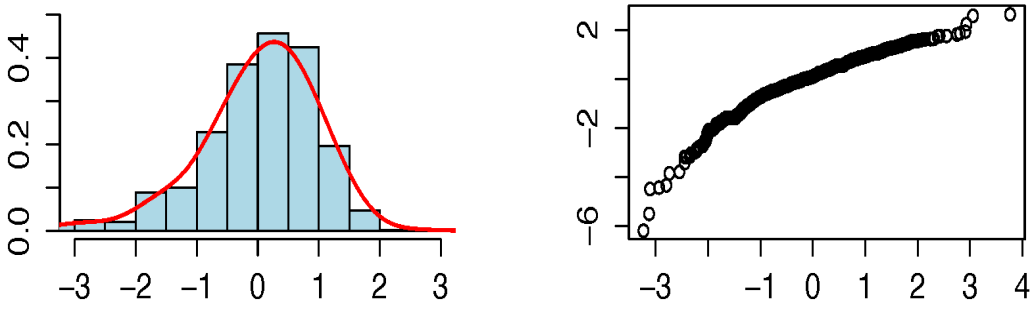

Figure 8.2: Histogram and qq-plots of d_2, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

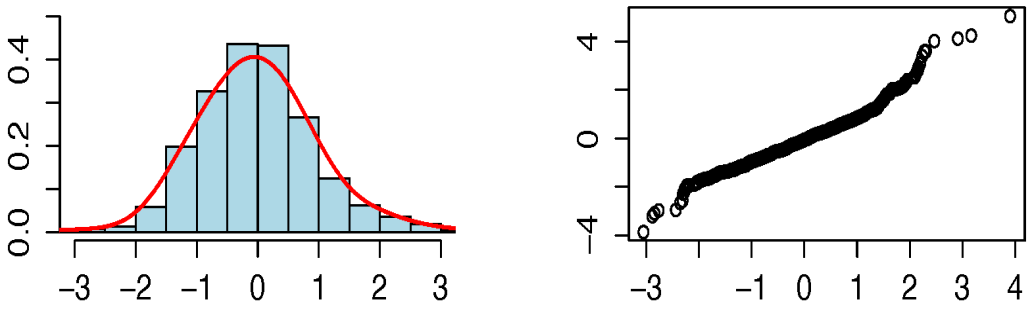

Figure 8.3: Histogram and qq-plots of a_{11}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

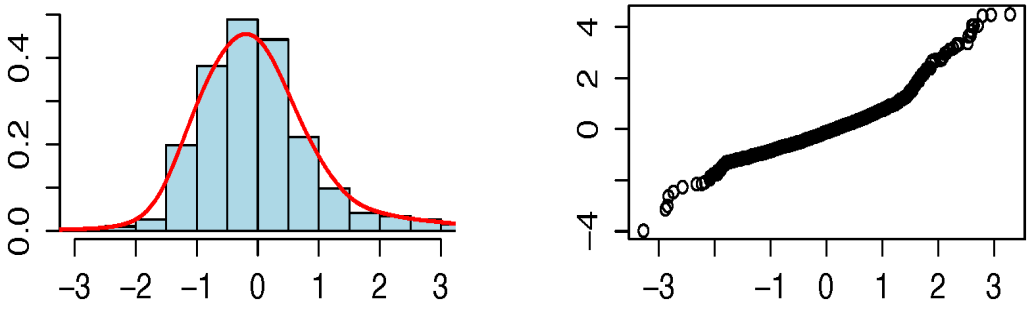

Figure 8.4: Histogram and qq-plots of a_{12}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

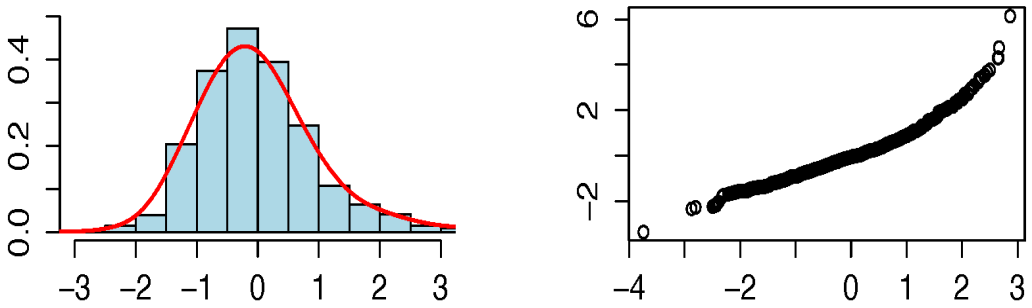

Figure 8.5: Histogram and qq-plots of a_{21}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

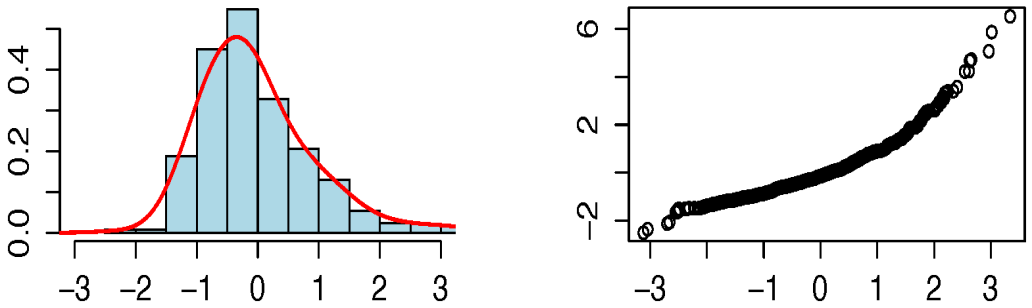

Figure 8.6: Histogram and qq-plots of a_{22}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

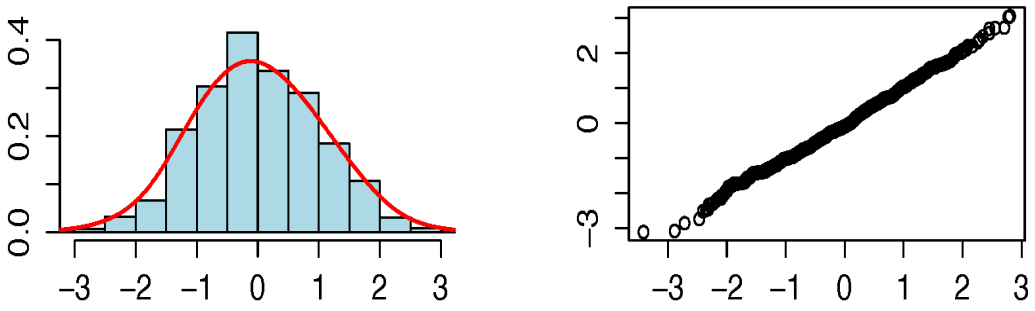

Figure 8.7: Histogram and qq-plots of b_{11}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

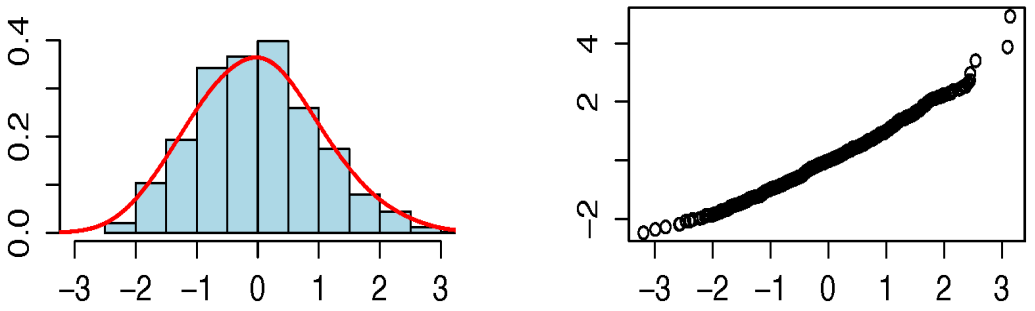

Figure 8.8: Histogram and qq-plots of b_{12}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

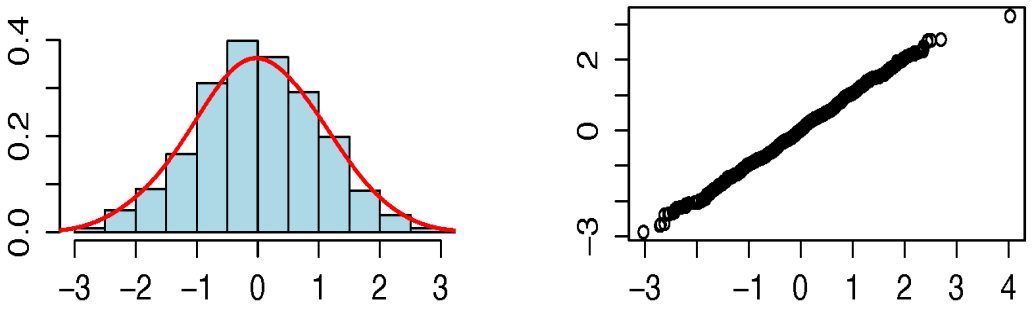

Figure 8.9: Histogram and qq-plots of b_{21}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

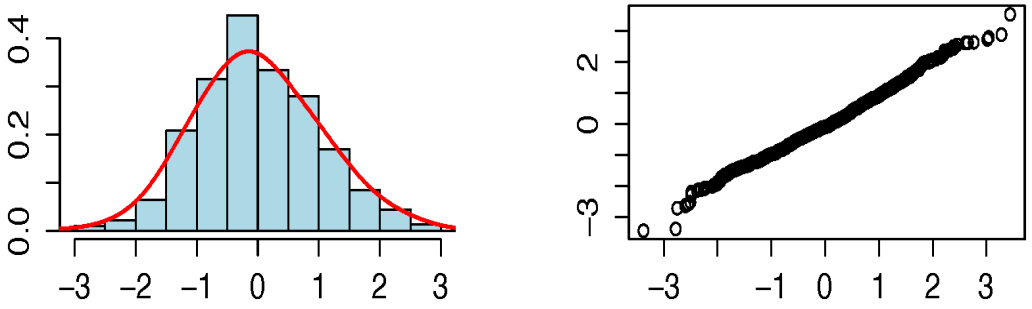

Figure 8.10: Histogram and qq-plots of b_{22}, for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

Table 8.2: Simulation Results for Bivariate Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (0.9, 0.8); eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.82, -0.32)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
200	d_1	0.3	0.899	0.018	0.617	2.246	8.25E-10
	d_2	0.7	0.681	0.024	1.202	9.480	4.16E-08
	a_{11}	0.7	0.203	0.016	-0.544	2.194	1.11E-05
	a_{12}	0.2	0.137	0.005	-0.574	3.712	0.002
	a_{21}	-0.1	-0.076	0.021	-1.110	8.594	2.13E-05
	a_{22}	0.4	0.386	0.005	-1.948	10.639	1.42E-06
	b_{11}	0.1	0.091	0.003	-0.018	3.166	0.573
	b_{12}	-0.2	-0.201	0.003	0.592	4.374	0.069
	b_{21}	-0.6	-0.596	0.004	0.960	6.640	0.341
	b_{22}	0.5	0.481	0.004	-0.611	5.072	0.078
500	d_1	0.3	0.518	0.010	2.106	8.544	1.70E-14
	d_2	0.7	0.650	0.012	-0.083	7.686	1.11E-05
	a_{11}	0.7	0.524	0.009	-2.059	8.184	3.48E-14
	a_{12}	0.2	0.195	0.002	-2.076	9.463	7.36E-11
	a_{21}	-0.1	-0.050	0.010	0.117	7.359	4.22E-04
	a_{22}	0.4	0.409	0.002	-0.473	5.015	0.148
	b_{11}	0.1	0.095	0.002	0.163	3.376	0.288
	b_{12}	-0.2	-0.213	0.002	0.357	3.505	0.181
	b_{21}	-0.6	-0.602	0.002	0.117	2.918	0.610
	b_{22}	0.5	0.488	0.002	0.134	3.174	0.828
1000	d_1	0.3	0.399	0.008	4.083	24.616	2.2E-16
	d_2	0.7	0.660	0.007	-1.094	10.493	8.67E-04
	a_{11}	0.7	0.620	0.007	-4.100	24.842	4.95E-14
	a_{12}	0.2	0.199	0.002	-3.322	21.277	1.12E-06
	a_{21}	-0.1	-0.061	0.006	1.153	10.585	0.020
	a_{22}	0.4	0.404	0.002	-6.802	76.067	1.86E-08
	b_{11}	0.1	0.097	0.001	0.093	2.750	0.914
	b_{12}	-0.2	-0.204	0.001	3.293	28.998	1.64E-04
	b_{21}	-0.6	-0.599	0.002	4.566	45.797	2.13E-05
	b_{22}	0.5	0.491	0.002	-2.729	24.557	0.006

Table 8.3: Simulation Results for Bivariate Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (-0.55, 0.3), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (-0.8, 0.65)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
200	d_1	0.6	0.711	0.009	3.712	30.349	2.43E-08
	d_2	0.4	0.511	0.011	1.110	6.615	0.020
	a_{11}	0.35	0.420	0.011	1.149	18.459	0.000
	a_{12}	0.1	-0.119	0.016	-3.224	27.053	1.12E-06
	a_{21}	0.7	0.672	0.012	-2.963	57.932	0.000
	a_{22}	-0.4	-0.506	0.011	-0.986	13.942	0.164
	b_{11}	-0.5	-0.506	0.003	1.017	6.338	0.263
	b_{12}	0.2	0.212	0.003	0.184	3.394	0.536
	b_{21}	-0.1	-0.096	0.003	0.297	2.998	0.370
	b_{22}	0.3	0.286	0.003	-0.061	2.997	0.536
500	d_1	0.6	0.615	0.003	0.185	3.451	0.648
	d_2	0.4	0.411	0.006	0.277	3.419	0.038
	a_{11}	0.35	0.379	0.005	0.884	5.838	0.078
	a_{12}	0.1	0.055	0.007	-1.040	6.188	0.181
	a_{21}	0.7	0.704	0.006	0.734	4.508	0.010
	a_{22}	-0.4	-0.420	0.006	-0.185	4.163	0.241
	b_{11}	-0.5	-0.505	0.002	0.044	3.076	0.536
	b_{12}	0.2	0.202	0.002	0.033	2.970	0.219
	b_{21}	-0.1	-0.099	0.002	0.088	2.936	0.685
	b_{22}	0.3	0.301	0.002	0.068	2.912	0.859
1000	d_1	0.6	0.610	0.002	0.115	3.240	0.432
	d_2	0.4	0.400	0.004	0.075	3.062	3.18E-08
	a_{11}	0.35	0.357	0.003	0.436	4.929	0.341
	a_{12}	0.1	0.083	0.004	-0.663	5.797	0.723
	a_{21}	0.7	0.697	0.004	0.461	4.018	0.314
	a_{22}	-0.4	-0.403	0.004	-0.224	5.735	0.370
	b_{11}	-0.5	-0.503	0.001	-0.091	2.835	0.401
	b_{12}	0.2	0.200	0.001	0.042	3.194	0.685
	b_{21}	-0.1	-0.099	0.001	0.051	2.867	0.573
	b_{22}	0.3	0.301	0.001	-0.187	2.997	0.466

Table 8.4: Simulation Results for Bivariate Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (0.7, 0.3), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.75, 0.75)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
200	d_1	0.6	0.628	0.008	0.740	4.738	2.20E-16
	d_2	0.2	0.289	0.008	1.433	8.338	1.38E-05
	a_{11}	-0.2	-0.213	0.007	0.350	3.342	0.263
	a_{12}	0.3	0.301	0.009	-5.684	94.423	1.07E-08
	a_{21}	-0.45	-0.529	0.009	-0.960	6.300	0.288
	a_{22}	0.1	0.052	0.008	-0.523	4.323	0.002
	b_{11}	0.5	0.480	0.003	0.015	2.989	0.980
	b_{12}	-0.3	-0.303	0.003	0.075	2.840	0.723
	b_{21}	-0.1	-0.101	0.004	-0.106	3.010	0.936
	b_{22}	0.6	0.564	0.004	-0.178	3.493	0.828
500	d_1	0.6	0.598	0.004	0.285	5.222	0.432
	d_2	0.2	0.235	0.005	0.651	3.920	0.048
	a_{11}	-0.2	-0.194	0.004	0.440	4.946	0.020
	a_{12}	0.3	0.317	0.003	-1.063	13.895	0.263
	a_{21}	-0.45	-0.482	0.005	-0.209	3.532	0.097
	a_{22}	0.1	0.086	0.004	-0.308	3.692	0.969
	b_{11}	0.5	0.493	0.002	-0.033	3.102	0.500
	b_{12}	-0.3	-0.300	0.002	0.067	3.231	0.241
	b_{21}	-0.1	-0.097	0.003	0.062	3.222	0.341
	b_{22}	0.6	0.585	0.003	-0.012	3.270	0.500
1000	d_1	0.6	0.596	0.003	-0.110	2.816	0.794
	d_2	0.2	0.219	0.003	0.298	3.032	0.087
	a_{11}	-0.2	-0.194	0.003	0.465	3.797	0.029
	a_{12}	0.3	0.305	0.002	0.105	2.927	0.759
	a_{21}	-0.45	-0.468	0.004	-0.130	3.048	0.828
	a_{22}	0.1	0.092	0.003	-0.174	3.181	0.314
	b_{11}	0.5	0.498	0.001	0.208	2.886	0.062
	b_{12}	-0.3	-0.301	0.001	-0.067	2.913	0.685
	b_{21}	-0.1	-0.101	0.002	0.179	3.380	0.097
	b_{22}	0.6	0.595	0.002	0.103	2.992	0.370

Table 8.5: Simulation Results for Bivariate Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (0.81, -0.51), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.66, -0.36)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
200	d_1	0.4	0.406	0.005	-0.011	3.739	1.35E-04
	d_2	0.7	0.673	0.005	0.293	11.464	0.401
	a_{11}	0.3	0.228	0.012	-1.222	5.763	5.07E-04
	a_{12}	-0.1	-0.106	0.010	-1.221	11.540	0.048
	a_{21}	-0.15	-0.079	0.014	-1.948	51.634	7.43E-05
	a_{22}	-0.3	-0.278	0.008	0.865	7.406	0.003
	b_{11}	0.1	0.095	0.004	0.349	3.328	0.573
	b_{12}	-0.4	-0.394	0.004	0.223	3.141	0.759
	b_{21}	-0.6	-0.593	0.004	0.133	3.001	0.432
	b_{22}	0.2	0.192	0.004	0.000	2.899	0.500
500	d_1	0.4	0.398	0.003	-0.254	3.125	0.003
	d_2	0.7	0.692	0.003	-0.194	3.532	0.432
	a_{11}	0.3	0.293	0.006	-0.649	4.641	0.013
	a_{12}	-0.1	-0.093	0.006	-0.217	3.182	0.610
	a_{21}	-0.15	-0.126	0.006	-0.190	3.874	0.888
	a_{22}	-0.3	-0.295	0.004	0.054	3.649	0.341
	b_{11}	0.1	0.097	0.003	0.017	3.200	0.200
	b_{12}	-0.4	-0.401	0.003	0.101	3.032	0.969
	b_{21}	-0.6	-0.597	0.003	0.012	3.280	0.969
	b_{22}	0.2	0.194	0.003	-0.023	3.109	0.723
1000	d_1	0.4	0.404	0.002	0.276	3.272	0.002
	d_2	0.7	0.692	0.002	0.145	3.369	0.500
	a_{11}	0.3	0.296	0.004	0.151	3.855	0.723
	a_{12}	-0.1	-0.101	0.004	-0.199	3.283	0.401
	a_{21}	-0.15	-0.135	0.004	0.097	3.289	0.888
	a_{22}	-0.3	-0.291	0.003	-0.135	3.276	0.263
	b_{11}	0.1	0.097	0.002	0.102	2.962	0.954
	b_{12}	-0.4	-0.402	0.002	0.146	2.759	0.828
	b_{21}	-0.6	-0.600	0.002	-0.096	2.873	0.573
	b_{22}	0.2	0.198	0.002	-0.050	2.914	0.648

Table 8.6: Simulation Results for Bivariate Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (-0.7, -0.3), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (-0.9, 0.9)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
200	d_1	0.9	0.902	0.012	1.030	6.682	0.022
	d_2	0.6	0.577	0.009	0.369	3.518	0.010
	a_{11}	-0.1	-0.105	0.007	-0.439	4.271	0.314
	a_{12}	-0.4	-0.394	0.013	-1.620	12.263	0.008
	a_{21}	-0.3	-0.276	0.006	-0.308	4.705	0.500
	a_{22}	-0.4	-0.370	0.008	0.560	3.719	0.015
	b_{11}	-0.3	-0.291	0.003	0.321	3.427	0.610
	b_{12}	0.7	0.684	0.003	0.016	2.916	0.097
	b_{21}	0.4	0.385	0.003	-0.025	3.119	0.164
	b_{22}	-0.2	-0.194	0.003	0.050	2.725	0.219
500	d_1	0.9	0.887	0.007	1.225	7.486	0.015
	d_2	0.6	0.572	0.006	0.071	3.737	0.022
	a_{11}	-0.1	-0.096	0.004	-0.914	2.849	0.148
	a_{12}	-0.4	-0.393	0.007	-1.265	8.131	0.108
	a_{21}	-0.3	-0.282	0.004	0.246	3.895	0.536
	a_{22}	-0.4	-0.382	0.005	0.596	4.150	0.134
	b_{11}	-0.3	-0.297	0.002	0.073	3.180	0.888
	b_{12}	0.7	0.702	0.002	0.103	3.275	0.828
	b_{21}	0.4	0.398	0.002	0.009	3.109	0.794
	b_{22}	-0.2	-0.195	0.002	-0.069	2.849	0.370
1000	d_1	0.9	0.894	0.004	-0.104	4.888	0.401
	d_2	0.6	0.584	0.004	0.417	4.676	0.314
	a_{11}	-0.1	-0.095	0.003	0.561	8.309	0.341
	a_{12}	-0.4	-0.400	0.004	0.177	7.103	0.888
	a_{21}	-0.3	-0.295	0.003	-0.200	5.636	0.466
	a_{22}	-0.4	-0.390	0.003	0.328	3.772	0.648
	b_{11}	-0.3	-0.300	0.001	0.034	3.091	0.988
	b_{12}	0.7	0.702	0.001	0.029	3.157	0.573
	b_{21}	0.4	0.404	0.002	0.008	2.774	0.969
	b_{22}	-0.2	-0.197	0.001	-0.066	3.182	0.500

8.2 Simulations for Three Dimensional Log-linear Poisson Autoregression

In this section, we expend the simulation study to three-dimensional model. Similar as bivariate model 6.1.3, the three-dimensional log-Poisson autoregression is expressed as

$$\begin{pmatrix} v_{t,1} \\ v_{t,2} \\ v_{t,3} \end{pmatrix} = \boldsymbol{D} + \boldsymbol{A} \begin{pmatrix} v_{t-1,1} \\ v_{t-1,2} \\ v_{t-1,3} \end{pmatrix} + \boldsymbol{B} \begin{pmatrix} log(Y_{t-1,1}+1) \\ log(Y_{t-1,2}+1) \\ log(Y_{t-1,3}+1) \end{pmatrix} \tag{8.2.1}$$

where $\boldsymbol{D} \in \mathbb{R}^3_+$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are 3×3 matrix with entries $\in \mathbb{R}$ restricted by the parameter space. The model 8.2.1 captures the dependence between $\{v_{t,1}\}$, $\{v_{t,2}\}$ and $\{v_{t,3}\}$. Follow the likelihood method stated in section 7.1, there are 21 parameters need to be estimated. The simulation for three-dimensional log-Poisson model are based on the sample size of 3000, and similar as the bivariate simulations, all the simulation studies are based on 1000 times. We also use a table to show the results for the estimations. Table 8.7 shows the results for the model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (-0.95, 0.34 and -0.21), and eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.87, 0.7, -0.24), with the same indicators we have considered in Bivariate simulations. And the histogram and qq-plots for each of the parameters are given in figure 8.11, figure 8.12 , figure 8.13, figure 8.14, figure 8.15, figure 8.16, figure 8.17. Also, more simulation results with different eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$) and ($\boldsymbol{A}$-$\boldsymbol{B}$) are given in table 8.8 and table 8.9.

Table 8.7: Simulation Results for 3-D Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (-0.95, 0.34 and -0.21), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.87, 0.7, -0.24)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
3000	d_1	0.7	0.690	0.003	1.454	12.601	5.47E-02
	d_2	0.3	0.299	0.003	-0.154	3.210	7.76E-02
	d_3	0.8	0.814	0.004	6.832	79.974	3.28E-07
	a_{11}	-0.50	-0.497	0.002	-1.462	17.554	2.00E-01
	a_{12}	-0.20	-0.198	0.001	-0.127	3.813	9.80E-01
	a_{13}	0.40	0.401	0.001	-1.628	13.549	9.71E-02
	a_{21}	0.20	0.202	0.002	2.071	19.882	2.19E-01
	a_{22}	-0.30	-0.298	0.001	0.881	7.689	6.16E-02
	a_{23}	-0.10	-0.101	0.001	0.100	3.652	6.85E-01
	a_{31}	-0.40	-0.408	0.003	-9.297	116.455	2.20E-16
	a_{32}	0.00	-0.003	0.002	-6.900	80.363	1.52E-11
	a_{33}	-0.28	-0.282	0.001	-0.140	15.167	2.57E-02
	b_{11}	0.20	0.201	0.001	-2.487	27.309	1.64E-01
	b_{12}	0.30	0.301	0.001	-0.561	5.979	4.66E-01
	b_{13}	-0.40	-0.399	0.001	0.239	4.360	3.41E-01
	b_{21}	0.10	0.100	0.001	-0.416	5.221	5.73E-01
	b_{22}	-0.30	-0.301	0.001	-2.068	28.889	1.64E-01
	b_{23}	0.50	0.500	0.001	0.034	3.456	7.59E-01
	b_{31}	-0.20	-0.201	0.001	-1.039	14.431	3.14E-01
	b_{32}	0.70	0.697	0.001	-10.615	235.160	3.51E-04
	b_{33}	0.35	0.349	0.001	0.247	3.736	5.36E-01

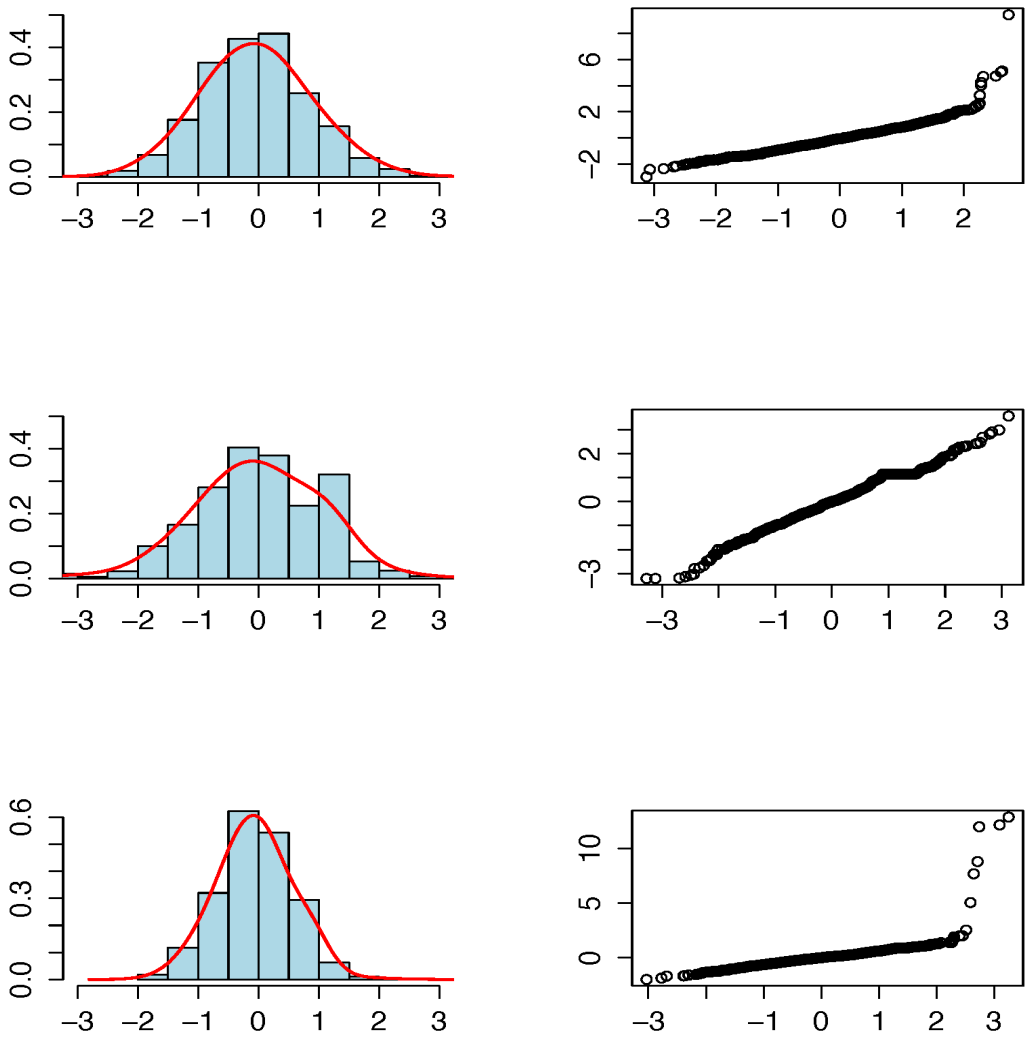

Figure 8.11: Histogram and qq-plots of d_1, d_2,d_3 (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

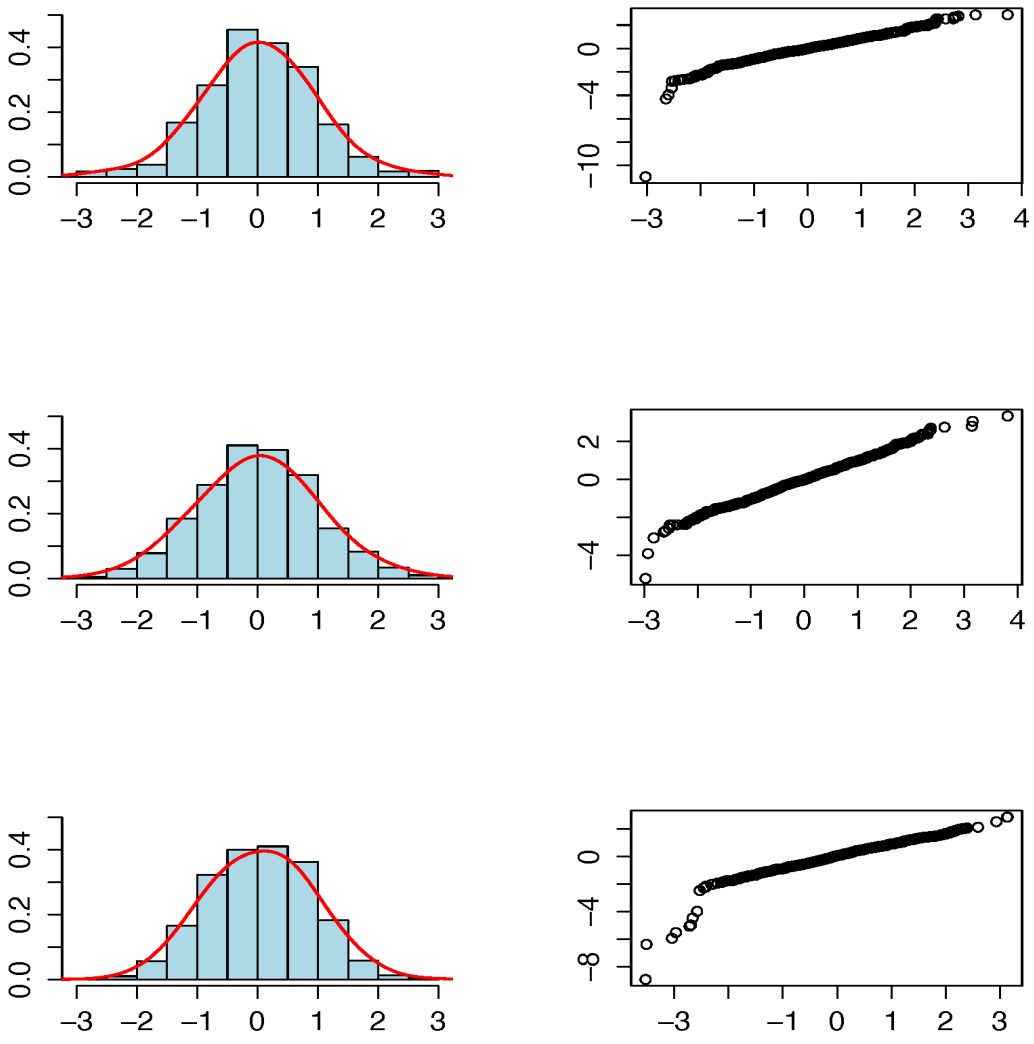

Figure 8.12: Histogram and qq-plots of a_{11}, a_{12}, a_{13} (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

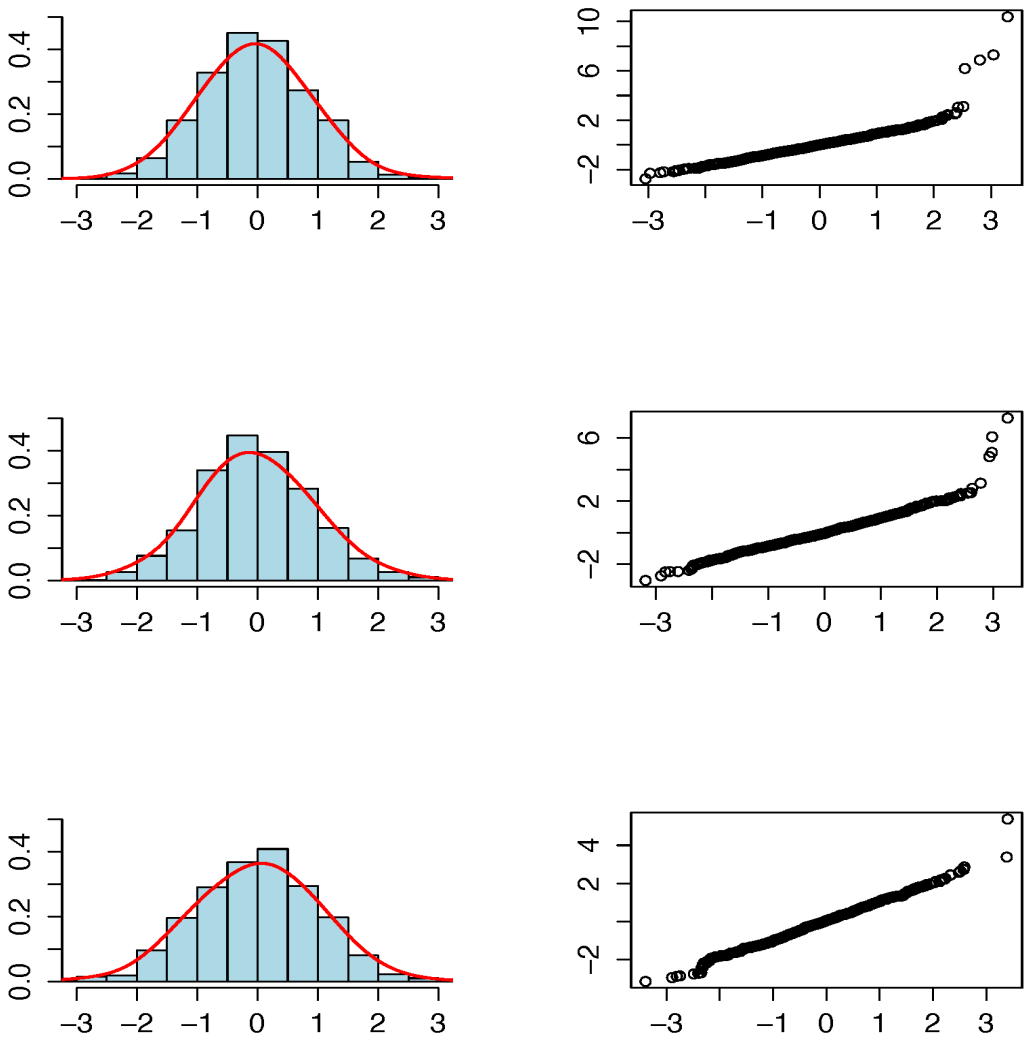

Figure 8.13: Histogram and qq-plots of a_{21}, a_{22}, a_{23} (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

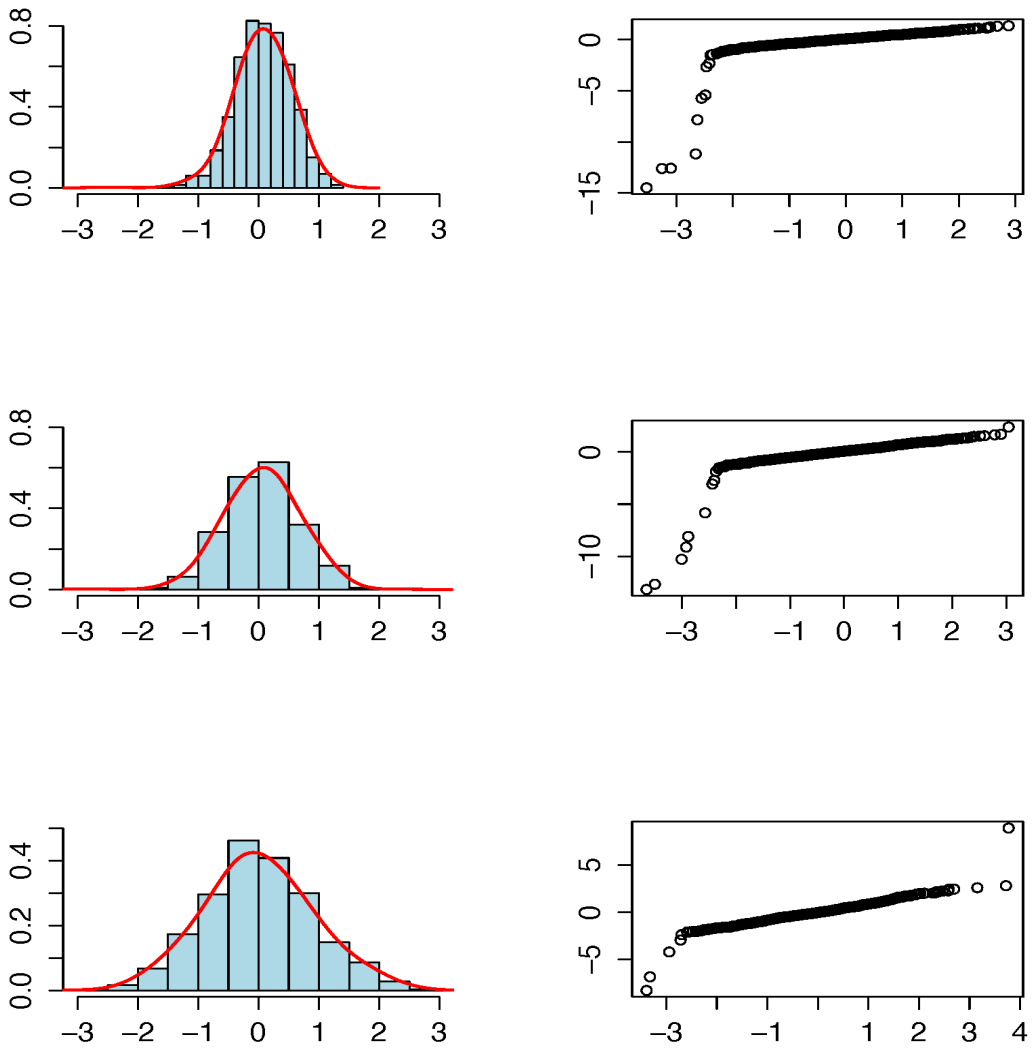

Figure 8.14: Histogram and qq-plots of a_{31}, a_{32}, a_{33} (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

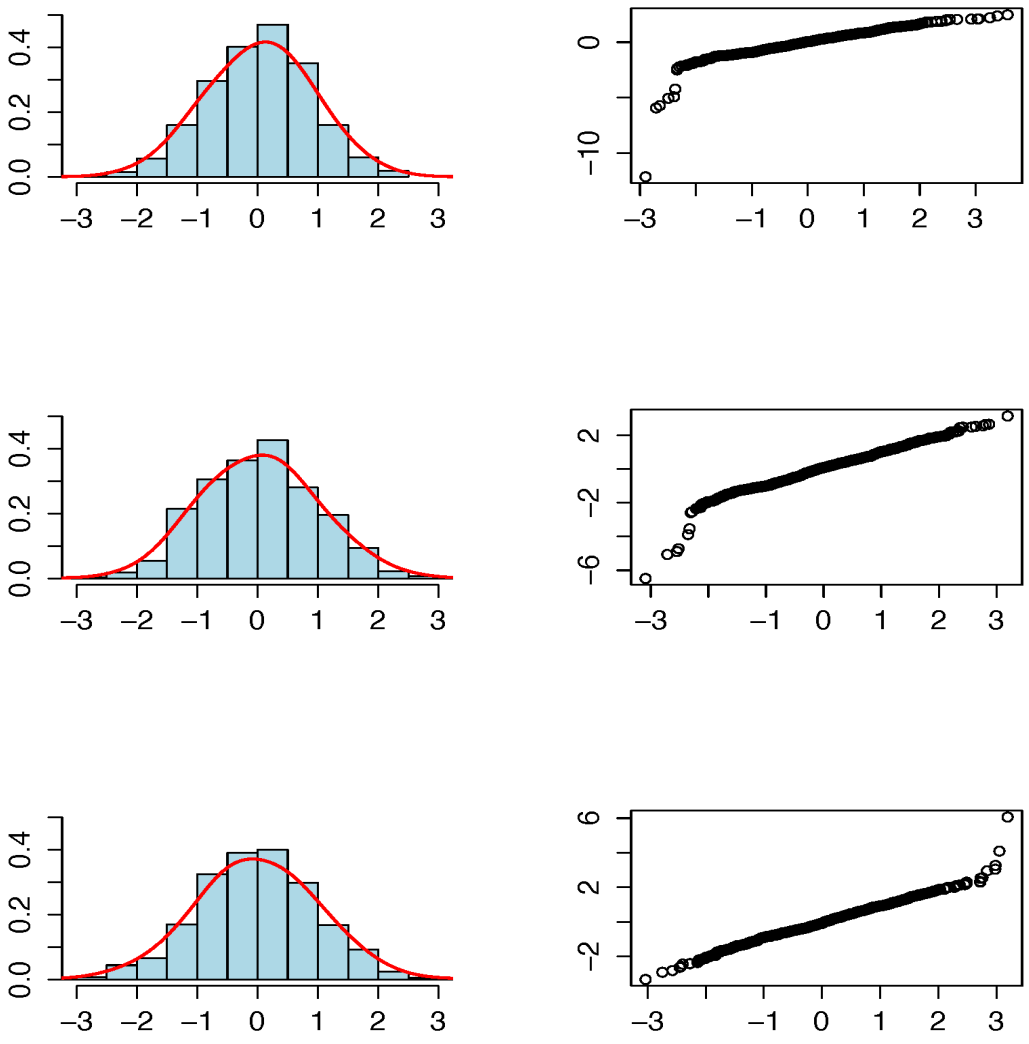

Figure 8.15: Histogram and qq-plots of b_{11}, b_{12}, b_{13} (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

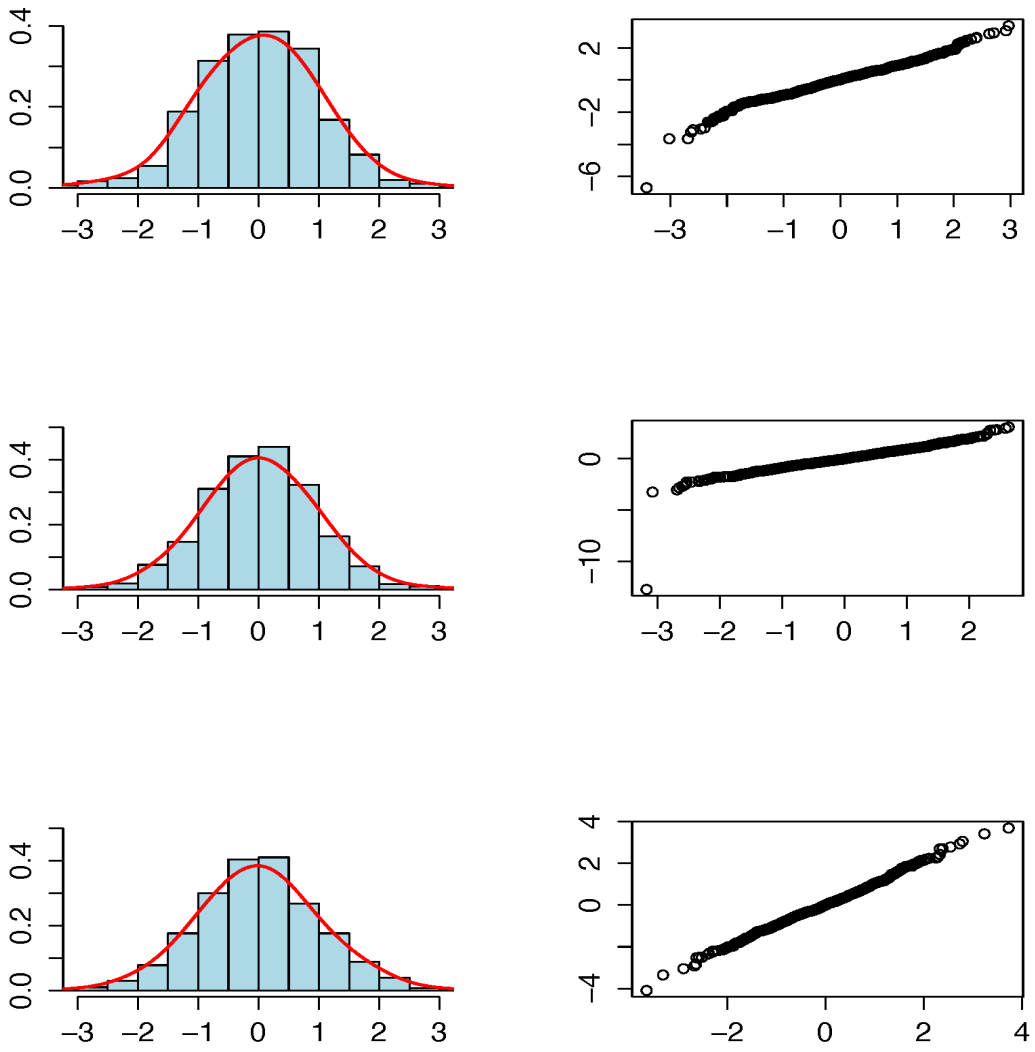

Figure 8.16: Histogram and qq-plots of b_{21}, b_{22}, b_{23} (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

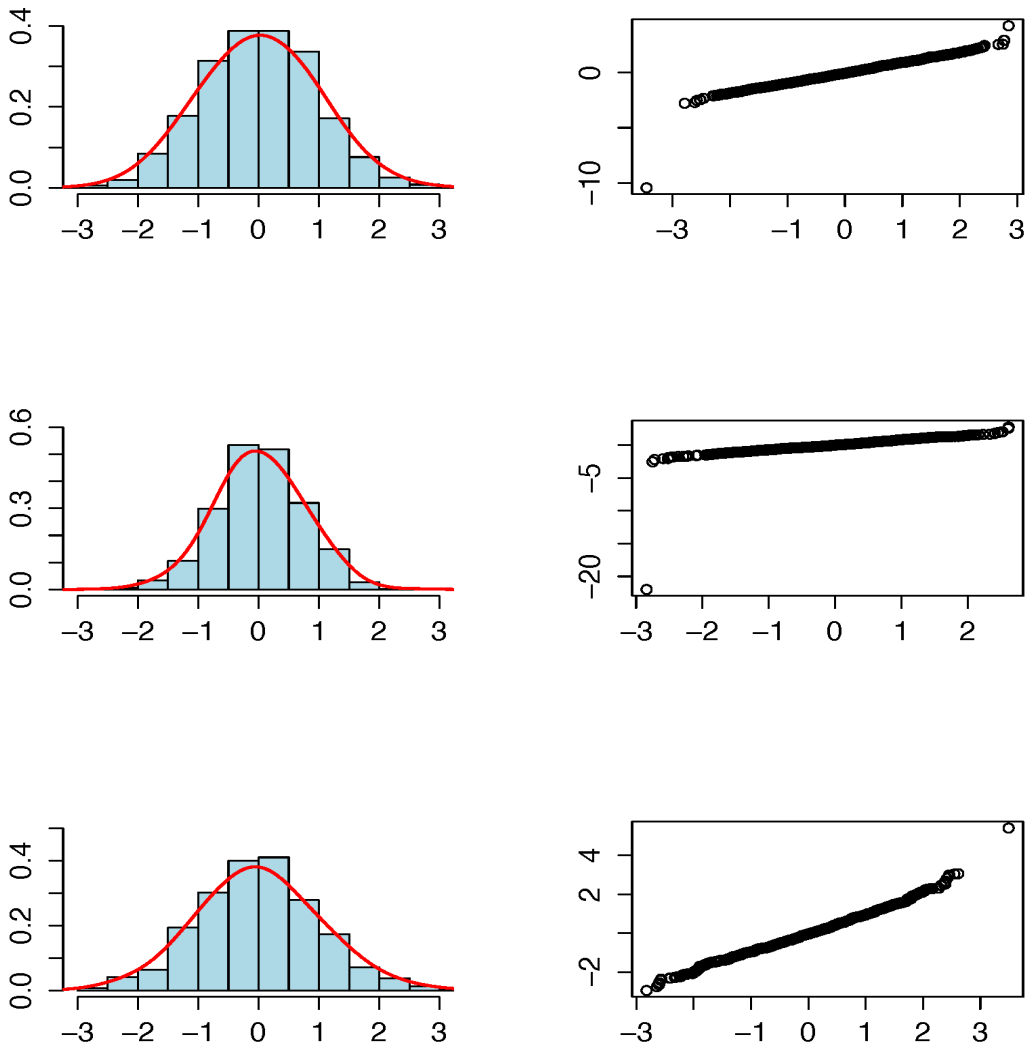

Figure 8.17: Histogram and qq-plots of b_{31}, b_{32}, b_{33} (from top to bottom), for the qq-lots, the horizontal axis is theoretical quantiles, the vertical axis is sample quantiles.

Table 8.8: Simulation Results for 3-D Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (0.82, -0.68 and 0.46), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.79, 0.79, 0.27)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
3000	d_1	0.5	0.591	0.014	8.320	89.431	2.20E-16
	d_2	0.3	0.341	0.006	4.860	40.409	2.20E-16
	d_3	0.8	0.870	0.014	6.840	65.276	2.20E-16
	a_{11}	0.25	0.153	0.012	-4.282	25.104	2.20E-16
	a_{12}	0.00	0.071	0.018	6.756	56.190	2.20E-16
	a_{13}	-0.30	-0.381	0.012	-6.763	63.474	2.20E-16
	a_{21}	0.10	0.060	0.007	-6.344	52.039	2.20E-16
	a_{22}	-0.30	-0.205	0.015	6.935	60.110	2.20E-16
	a_{23}	-0.20	-0.215	0.005	-6.592	63.778	2.20E-16
	a_{31}	0.40	0.243	0.032	-8.227	87.447	2.20E-16
	a_{32}	-0.50	-0.189	0.063	8.196	86.516	2.20E-16
	a_{33}	-0.30	-0.304	0.010	-6.718	90.641	2.20E-16
	b_{11}	0.50	0.497	0.002	-2.071	24.138	2.20E-16
	b_{12}	0.70	0.680	0.004	-4.261	25.926	2.20E-16
	b_{13}	-0.40	-0.388	0.002	4.603	29.645	2.20E-16
	b_{21}	0.60	0.569	0.004	-3.282	14.121	2.20E-16
	b_{22}	0.10	0.094	0.002	-1.981	13.278	2.20E-16
	b_{23}	-0.30	-0.284	0.002	2.600	11.329	2.20E-16
	b_{31}	0.20	0.178	0.003	-3.902	18.013	2.20E-16
	b_{32}	-0.10	-0.106	0.002	-5.721	46.957	2.20E-16
	b_{33}	0.35	0.337	0.002	-2.755	12.236	2.20E-16

Table 8.9: Simulation Results for 3-D Log-linear Poisson Autoregression Model with eigenvalue of ($\boldsymbol{A}$+$\boldsymbol{B}$): (0.77, -0.59 and 0.22), eigenvalue of ($\boldsymbol{A}$-$\boldsymbol{B}$): (0.61, 0.49, -0.32)

Sample Size	Parameters	Real Value	MLE	Standard Error	Skewness	Kurtosis	P-Value
3000	d_1	0.4	0.133	0.006	1.504	4.703	2.20E-16
	d_2	0.7	1.039	0.008	0.094	3.911	0.007
	d_3	0.8	0.934	0.008	-0.539	3.891	1.07E-08
	a_{11}	0.70	0.831	0.003	-1.220	4.227	6.15E-09
	a_{12}	-0.30	-0.254	0.002	0.025	3.373	0.610
	a_{13}	0.45	0.541	0.003	-0.432	3.498	0.241
	a_{21}	-0.40	-0.623	0.005	-0.291	3.975	0.314
	a_{22}	0.65	0.512	0.004	-0.386	3.650	0.263
	a_{23}	-0.30	-0.406	0.003	-0.302	4.114	0.121
	a_{31}	-0.10	-0.192	0.005	0.388	3.696	0.007
	a_{32}	0.70	0.631	0.004	0.346	3.926	0.029
	a_{33}	-0.50	-0.535	0.003	0.326	3.888	0.219
	b_{11}	-0.20	-0.182	0.001	0.041	3.109	0.794
	b_{12}	0.10	0.130	0.001	-0.519	3.455	0.097
	b_{13}	-0.45	-0.432	0.001	-0.008	3.338	9.995E-01
	b_{21}	-0.10	-0.103	0.001	-0.052	2.871	0.314
	b_{22}	-0.35	-0.355	0.001	-0.110	2.734	0.859
	b_{23}	0.50	0.498	0.001	-0.051	2.917	0.536
	b_{31}	0.40	0.402	0.001	-0.139	3.279	0.759
	b_{32}	0.00	0.003	0.001	-0.051	3.276	0.828
	b_{33}	0.10	0.101	0.001	0.102	3.016	0.288

Chapter 9

Real Data Analysis for Multivariate Log-Poisson Autogression

9.1 Data Description

In this part, we apply this model to the stock return data to study the interdependence of different stocks in the same industry. By focusing on the exceedance of different stock return in a certain time period, the correlations of the exceedance return for the certain stocks can be gained by this model. Especially, we apply this model to study the dependence of financial risks in different market. Specifically, consider the stock market indices in the US, Europe, and Japan, the exceedance of

the stock index return over certain threshold represents the magnitude of market variations and provides us a new measurement for the market tail risk in different countries/market. Also, another application is the interdependence of trading behavior for different stocks, through which the impact of one stockâĂŹs trading behavior on another stock can be quantitatively modeled and identified by this model. In this study, we apply this model to the stock data for the number of trades per minute or within two minutes or five minutes and get the interaction of different stocks' transactions.

The data we use in this real data study are the stock daily return data and the TAQ data of the high frequency transactions. For the stock daily return data, given a threshold of return, let the Poisson event be the exceedance return and count the number of this event in one month. 20 years' daily data are used between Jan 1996 and Dec 2015 and the total of 240 months are considered as the sample size in this study. Index data like SP500, NASDAQ, FTSE 100 in London stock exchange, and Nikkei 225 in Japan market, different industry data like Bank of America, Citi group as the banking industry, GE and Philips as the machine industry, Exxon Mobil corporation and Marathon oil as the petroleum industry are considered in this study. For the TAQ data, we choose stock data in Jan 2002 and Jan 2012 respectively for comparison.

9.2 Exceedance Return Examples

9.2.1 Application to the dependence of indices/stocks in the same market/industry

We first apply the bivariate log-Poisson autoregression model to the analysis of the dependence of two indices/stocks within the same market/industry in this part. 20 years' indices/stock return data between Jan 1996 and Dec 2015 are used to analyze the interdependence of two markets/stocks in the same industry. Figure 9.1, 9.2, 9.3, 9.4, respectively show the plots of daily return during this period for the index of SP500 and NASDAQ, and other six stocks (of which each two are in the same industry) that we pick as the examples of three different industries in this study.

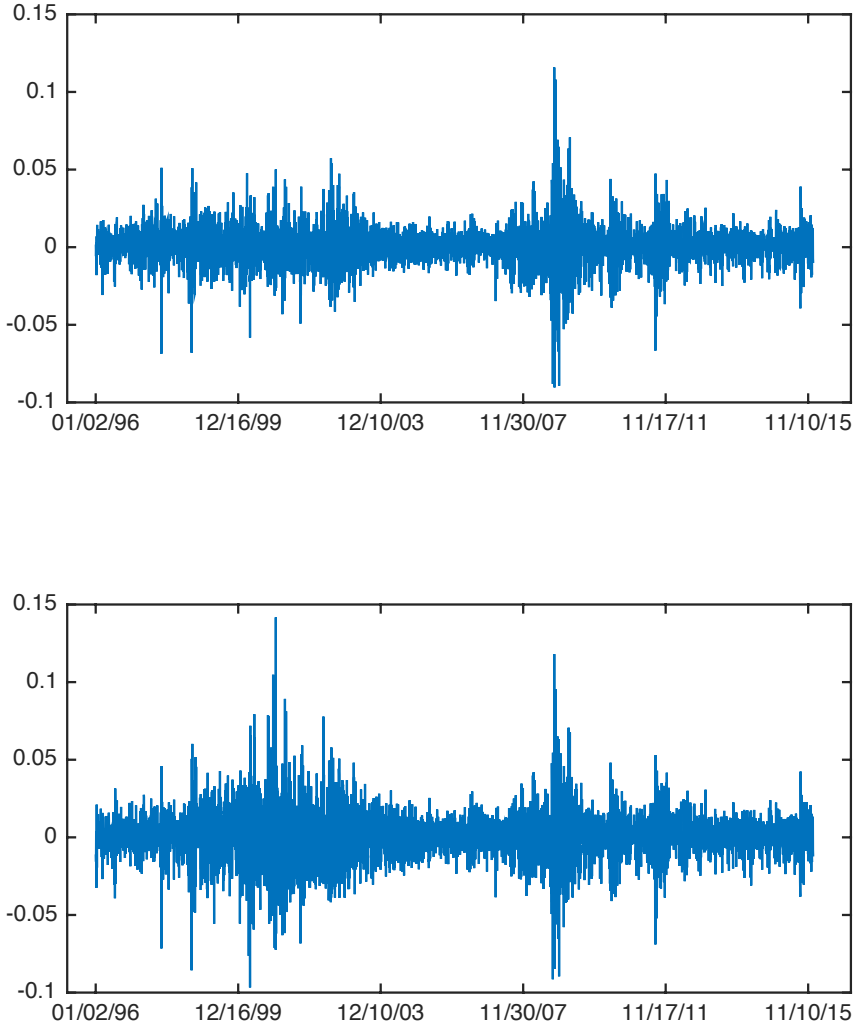

Figure 9.1: 20 Years daily return for SP500 and NASDAQ (from top to bottom), the horizontal axis is the time period from Jan 1996 to Dec 2015, the vertical axis is the daily return.

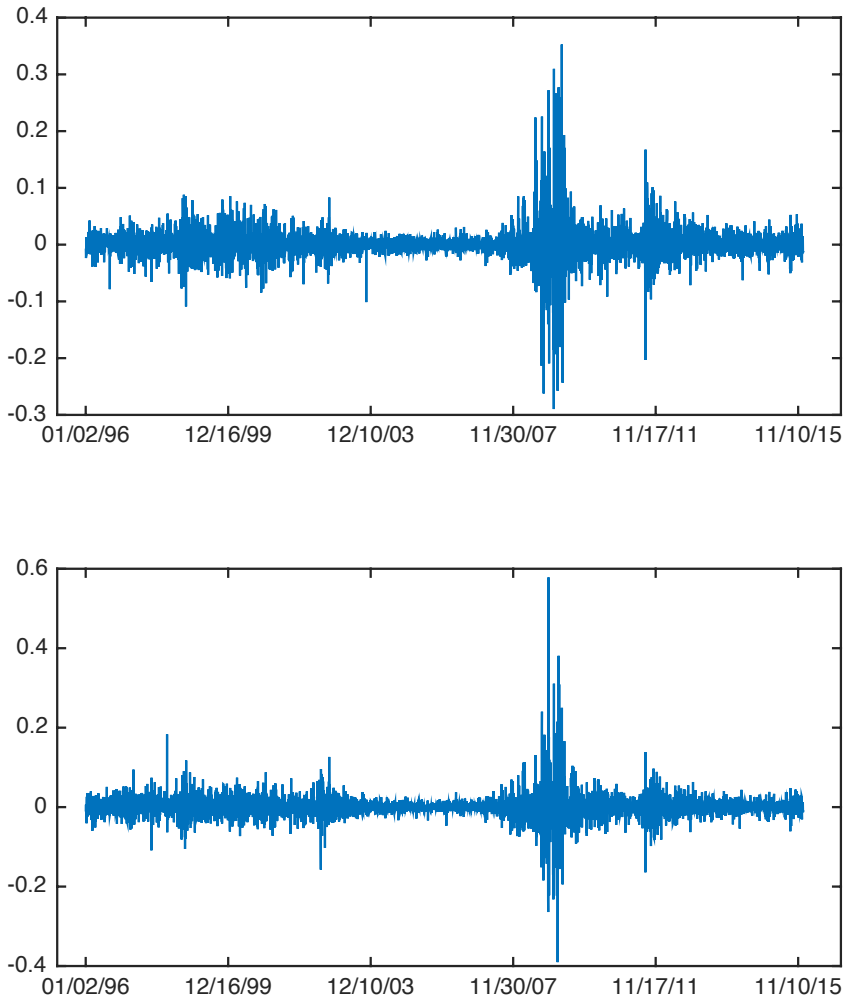

Figure 9.2: 20 Years daily return for BOA and Citigroup (from top to bottom), the horizontal axis is the time period from Jan 1996 to Dec 2015, the vertical axis is the daily return.

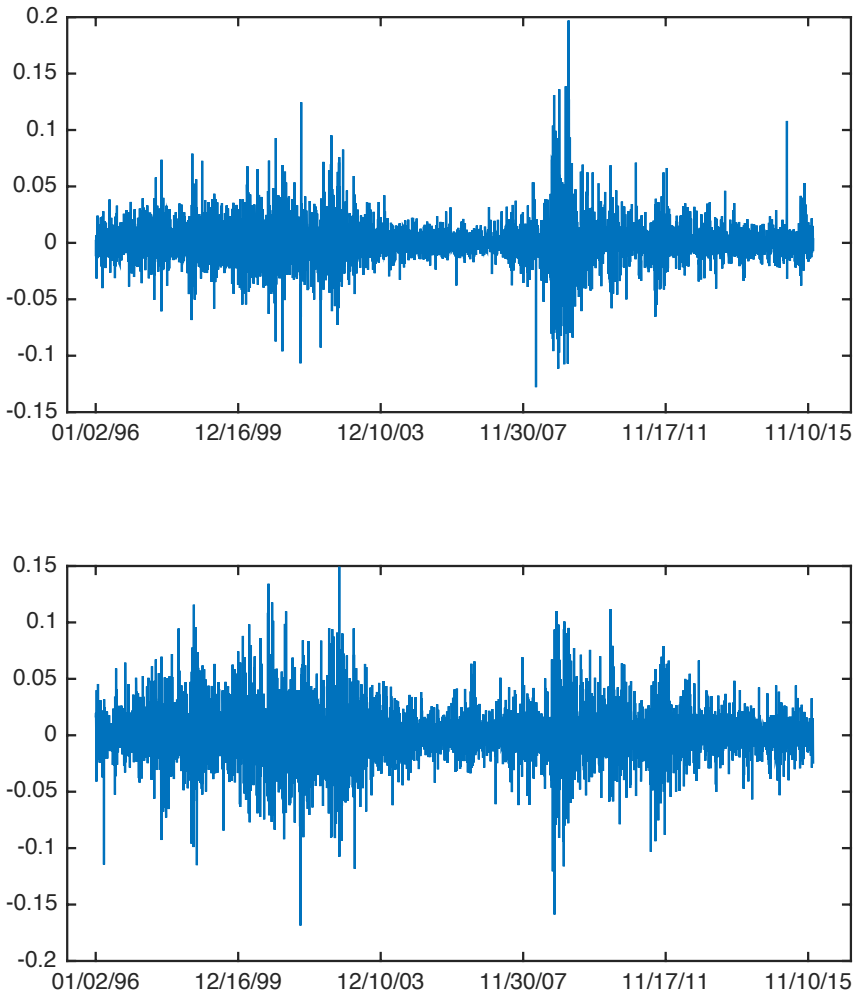

Figure 9.3: 20 Years daily return for GE and Philips (from top to bottom), the horizontal axis is the time period from Jan 1996 to Dec 2015, the vertical axis is the daily return.

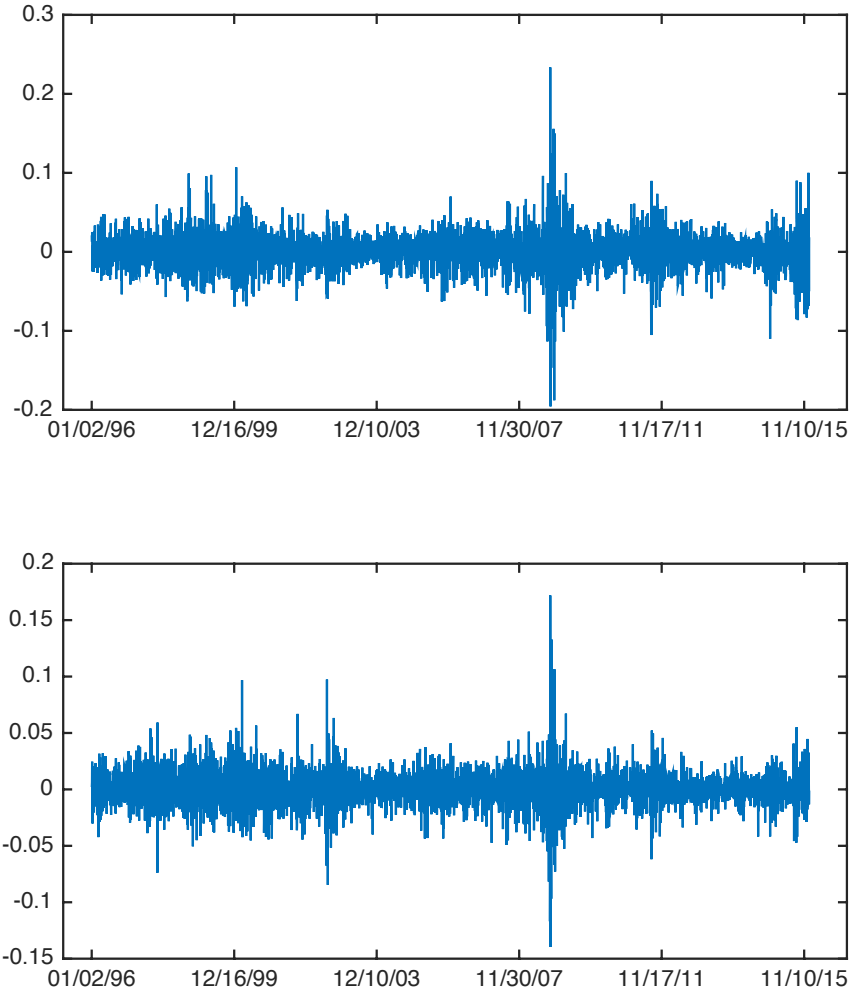

Figure 9.4: 20 Years daily return for Marathon Oil and Exxon Mobil (from top to bottom), the horizontal axis is the time period from Jan 1996 to Dec 2015, the vertical axis is the daily return.

Then, let the event be the number of exceedance of stock return per month given a specific threshold. There are total 240 months in our data set. Let $\{Y_{t,1}, Y_{t,2}\}$ be the bivariate observations, where $t = 1, 2, \ldots, 240$, $\{Y_{t,1}, t \geq 1\}$ and $\{Y_{t,2}, t \geq 1\}$ are two conditional independent time series of counts. We also plot the observations given different threshold at $(0.01, 0.01)$, $(-0.01, -0.02)$, $(0.02, -0.01)$ for SP500 v.s NASDAQ, at $(0.02, 0.02)$, $(-0.02, 0.05)$, $(0.01, -0.02)$ for Bank of America v.s Citigroup, at $(-0.03, -0.01)$, $(0.03, 0.02)$, $(0.01, 0.04)$ for GE v.s Philips, and at $(-0.01, -0.01)$, $(0.02, -0.03)$, $(0.03, 0.01)$ for Marathon Oil v.s Exxon Mobil. The corresponding log-Poisson autogregression and analysis are in the followings.

Given different thresholds, we plot the exceedance number per month for SP500 and NASDAQ respectively in figure 9.5. Let λ_1 represent the threshold of daily return for SP500, λ_2 represent the threshold of daily return for NASDAQ, the observations of SP500 and NASDAQ are $\{Y_{t,1}\}$ and $\{Y_{t,2}\}$ corresponding to model 6.1.3. Parameters can be estimated by direct optimization of the log-likelihood function 7.1.3 and the estimation results of the parameters in the bivariate log-Poisson autoregression model are shown in table 9.1 and 9.2 under different thresholds.

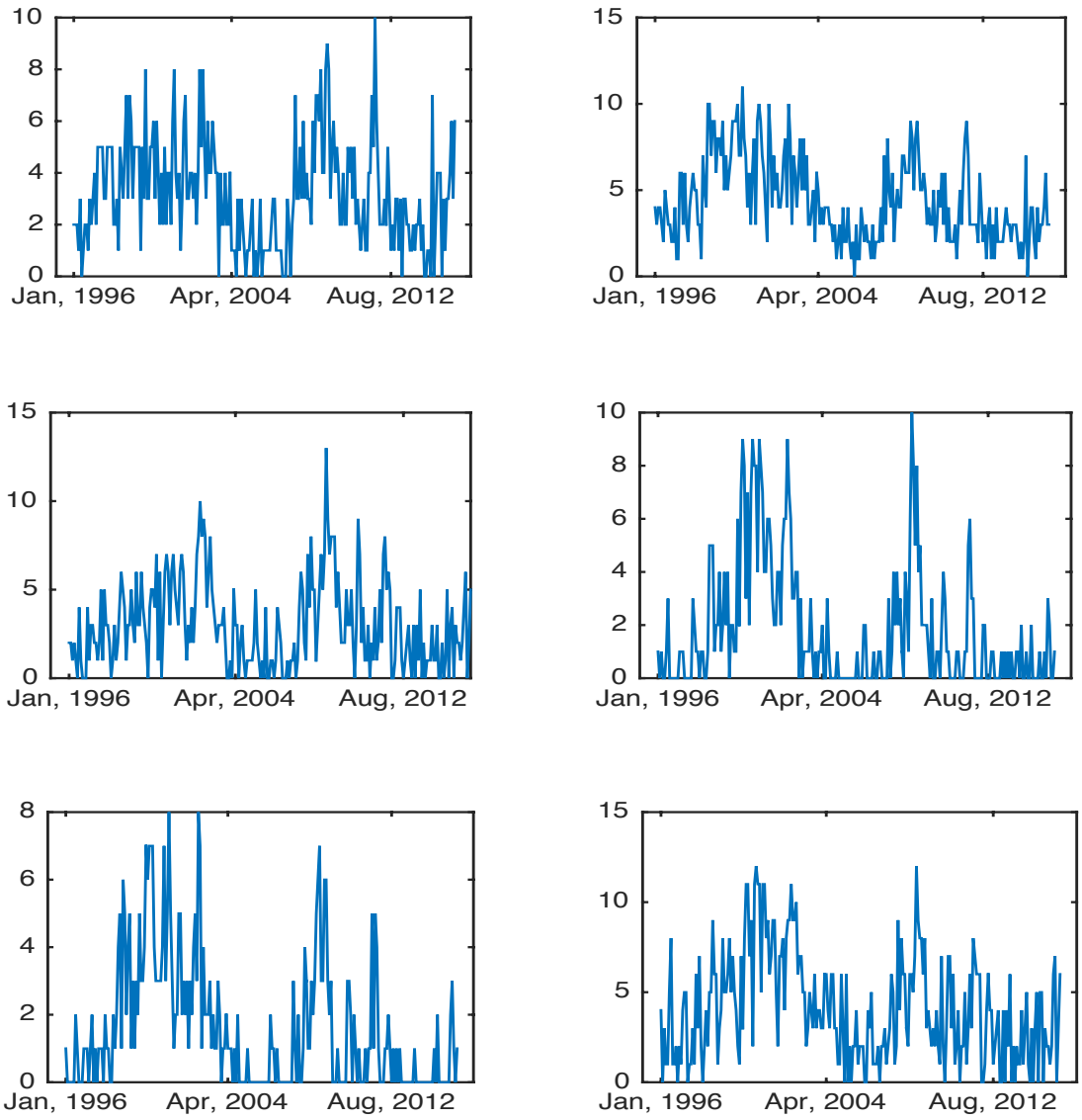

Figure 9.5: Exceedance number per month for SP500 (left) and NASDAQ (right) given different threshold (0.01, 0.01), (-0.01, -0.02), (0.02, -0.01) from top to bottom.

Table 9.1: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for SP500 and NASDAQ under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(0.01, 0.01)			$\boldsymbol{\lambda}$=(0.01, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.002	0.027	9.48E-01	1.096	0.489	2.60E-02
d_2	5.78E-09	0.016	1.00E+00	0.124	1.034	9.05E-01
a_{11}	0.650	0.055	1.13E-25	-0.305	0.487	5.32E-01
a_{12}	-0.028	0.057	6.24E-01	0.284	0.187	1.30E-01
a_{21}	-0.099	0.034	3.72E-03	-0.580	1.064	5.86E-01
a_{22}	0.725	0.035	1.27E-55	0.464	0.454	3.08E-01
b_{11}	0.319	0.038	2.17E-15	0.109	0.040	6.43E-03
b_{12}	0.009	0.039	8.12E-01	0.204	0.028	5.97E-12
b_{21}	0.140	0.021	1.04E-10	0.124	0.119	2.96E-01
b_{22}	0.204	0.024	1.25E-15	0.722	0.110	3.63E-10
	$\boldsymbol{\lambda}$=(0.01, -0.01)			$\boldsymbol{\lambda}$=(-0.01, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	2.31E-04	0.043	9.96E-01	1.83E-08	0.042	1.00E+00
d_2	0.405	0.089	8.28E-06	0.020	0.012	9.15E-02
a_{11}	0.824	0.112	2.97E-12	0.340	0.066	6.26E-07
a_{12}	-0.289	0.106	6.93E-03	0.169	0.063	7.56E-03
a_{21}	0.995	0.129	2.82E-13	-0.157	0.036	1.57E-05
a_{22}	-0.360	0.119	2.75E-03	0.839	0.029	3.10E-80
b_{11}	0.182	0.038	1.31E-07	0.408	0.046	2.82E-16
b_{12}	0.240	0.039	2.58E-25	-0.016	0.042	7.09E-01
b_{21}	-0.146	0.021	4.09E-12	0.077	0.021	2.79E-04
b_{22}	0.371	0.024	3.17E-45	0.188	0.019	9.26E-20
	$\boldsymbol{\lambda}$=(-0.02, -0.01)			$\boldsymbol{\lambda}$=(-0.01, -0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.094	0.364	7.96E-01	0.799	0.162	1.57E-06
d_2	0.546	0.142	1.59E-04	0.119	0.407	7.71E-01
a_{11}	0.614	0.166	2.71E-04	-0.215	0.180	2.33E-01
a_{12}	-0.797	0.401	4.82E-02	0.101	0.103	3.26E-01
a_{21}	0.118	0.057	3.91E-02	-0.446	0.520	3.91E-01
a_{22}	0.314	0.102	2.33E-03	0.402	0.270	1.38E-01
b_{11}	0.508	0.121	3.52E-05	0.112	0.037	2.74E-03
b_{12}	0.484	0.230	3.62E-02	0.435	0.034	7.60E-29
b_{21}	0.037	0.028	1.92E-01	-0.052	0.113	6.49E-01
b_{22}	0.291	0.032	6.40E-17	0.840	0.111	7.45E-13

Table 9.2: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for SP500 and NASDAQ under different threshold $\boldsymbol{\lambda}$ continued

	$\boldsymbol{\lambda}$=(0.02, -0.01)			$\boldsymbol{\lambda}$=(-0.01, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.10E-05	0.516	1.00E+00	0.231	0.134	8.65E-02
d_2	0.455	0.278	1.03E-01	0.003	0.245	9.92E-01
a_{11}	0.783	0.170	6.83E-06	0.342	0.148	2.13E-02
a_{12}	-0.793	0.495	1.11E-01	0.027	0.077	7.30E-01
a_{21}	0.064	0.091	4.83E-01	-0.832	0.307	7.20E-03
a_{22}	0.401	0.180	2.65E-02	0.733	0.148	1.34E-06
b_{11}	0.273	0.088	2.04E-03	0.313	0.034	7.71E-18
b_{12}	0.622	0.263	1.88E-02	0.105	0.036	3.40E-03
b_{21}	0.046	0.034	1.84E-01	0.543	0.102	2.13E-07
b_{22}	0.267	0.038	1.81E-11	0.378	0.075	8.03E-07
	$\boldsymbol{\lambda}$=(-0.02, 0.01)			$\boldsymbol{\lambda}$=(-0.01, 0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.023	0.603	9.69E-01	3.85E-09	0.093	1.00E+00
d_2	1.159	0.316	2.97E-04	2.64E-06	0.019	1.00E+00
a_{11}	0.624	0.224	5.78E-03	-0.064	0.086	4.53E-01
a_{12}	-0.811	0.348	2.07E-02	0.607	0.101	6.28E-09
a_{21}	0.285	0.109	9.38E-03	-0.098	0.047	3.72E-02
a_{22}	0.064	0.193	7.42E-01	0.652	0.049	1.22E-30
b_{11}	0.531	0.092	2.54E-08	0.434	0.032	7.24E-32
b_{12}	0.529	0.201	8.94E-03	-0.156	0.040	1.23E-04
b_{21}	0.166	0.020	1.26E-14	0.199	0.014	1.00E-32
b_{22}	0.116	0.037	1.92E-03	0.226	0.023	4.13E-19
	$\boldsymbol{\lambda}$=(0.02, 0.02)			$\boldsymbol{\lambda}$=(0.03, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	6.38E-10	0.053	1.00E+00	5.09E-10	0.164	1.00E+00
d_2	0.018	0.049	7.07E-01	0.046	0.067	4.92E-01
a_{11}	0.625	0.117	2.15E-07	0.399	0.391	3.08E-01
a_{12}	-0.426	0.126	8.23E-04	-0.548	0.584	3.50E-01
a_{21}	-0.216	0.079	7.00E-03	-0.153	0.584	1.97E-01
a_{22}	0.409	0.084	2.23E-06	0.244	0.116	3.70E-02
b_{11}	0.494	0.132	2.33E-04	0.800	0.586	1.74E-01
b_{12}	0.078	0.112	4.86E-01	-0.260	0.405	5.20E-01
b_{21}	0.322	0.058	6.12E-08	0.255	0.063	6.85E-05
b_{22}	0.277	0.070	9.77E-05	0.343	0.086	8.61E-05

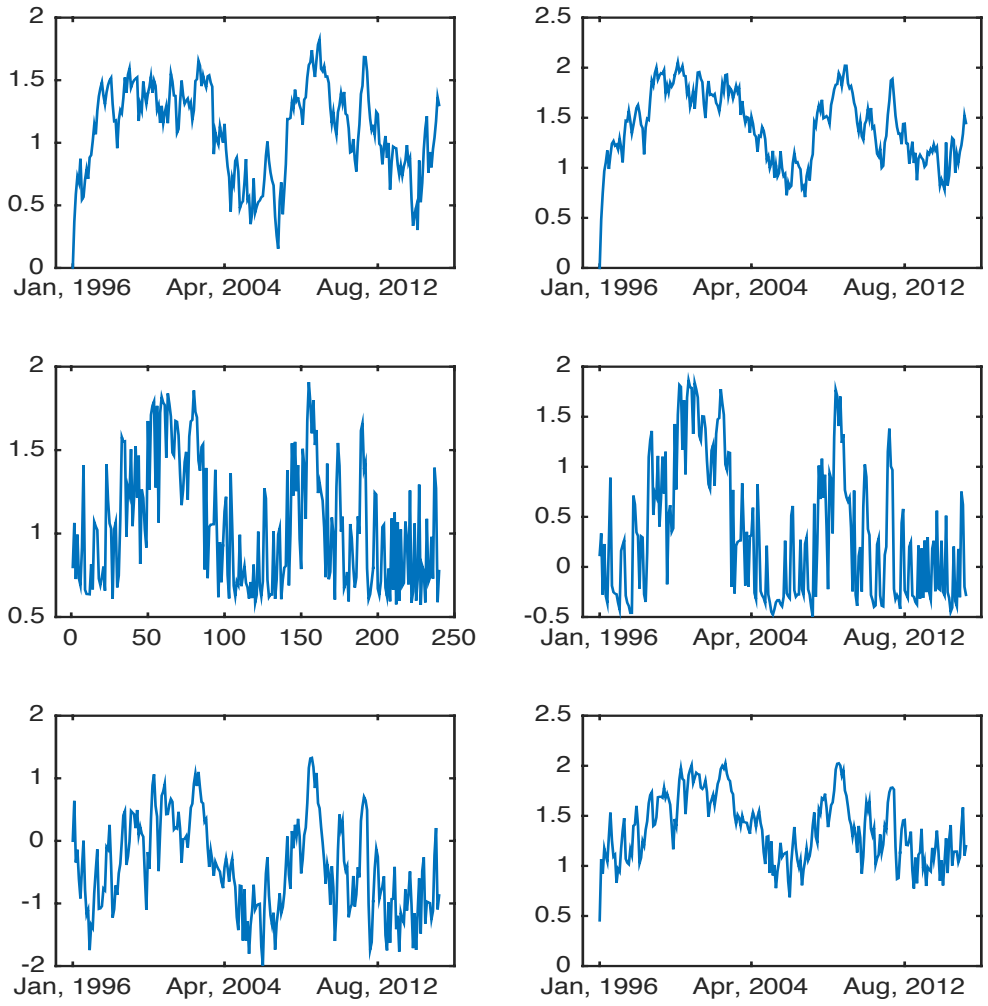

Figure 9.6: Time series of v_t in model 6.1.3 for SP500 (left) and NASDAQ (right) given different threshold (0.01, 0.01), (-0.01, -0.02), (0.02, -0.01) from top to bottom.

In the above tables, the second column and the fifth column list the estimations from the maximum likelihood method. The third column and the sixth column show the standard error of the corresponding estimated coefficients. The fourth column and the last column in the table are the P-value of t-test statistic for the estimators. In table 9.1, under the paired thresholds $\boldsymbol{\lambda} = (0.01, 0.01)$, from the estimated number of a_{12} and a_{21}, b_{12} and b_{21} in model 6.1.3, the interdependence between $v_{t+1,SP500}$ and $v_{t,NASDAQ}$ is -0.028, the interdependence between $v_{t+1,NASDAQ}$ and $v_{t,SP500}$ is -0.099, and the interdependence between $v_{t+1,SP500}$ and $log(Y_{t,NASDAQ}+1)$ is 0.622, the interdependence between $v_{t+1,NASDAQ}$ and $log(Y_{t,SP500}+1)$ is 0.046. While, with changing one of the paired thresholds, say $\boldsymbol{\lambda} = (0.01, 0.02)$, the interdependence between $v_{t+1,SP500}$ and $v_{t,NASDAQ}$ becomes positive at 0.284, the interdependence between $v_{t+1,NASDAQ}$ and $v_{t,SP500}$ is -0.580, and the interdependence between $v_{t+1,SP500}$ and $log(Y_{t,NASDAQ}+1)$ becomes a significant estimation at 0.204, the interdependence between $v_{t+1,NASDAQ}$ and $log(Y_{t,SP500}+1)$ becomes 0.124. When fixing the threshold for NASDAQ at -0.01, change the threshold of SP500 from 0.01 to -0.01, the interdependence between $v_{t+1,SP500}$ and $v_{t,NASDAQ}$ changes from -0.289 to 0.169, the interdependence between $v_{t+1,NASDAQ}$ and $v_{t,SP500}$ changes from 0.995 to -0.157, the interdependence between $v_{t+1,SP500}$ and $log(Y_{t,NASDAQ}+1)$ changes from 0.240 to -0.016, and the interdependence between $v_{t+1,NASDAQ}$ and $log(Y_{t,SP500}+1)$ changes from 0.037 to -0.052. Similar results can be seen from table 9.2. The insignificant estimation under the paired thresholds $\boldsymbol{\lambda} = (0.03, 0.02)$ in the right bottom corner in table 9.2 due to the reason of fewer observations under this threshold. Time series of $v_{t+1,1}$ and $v_{t+1,2}$ are plotted in figure 9.6.

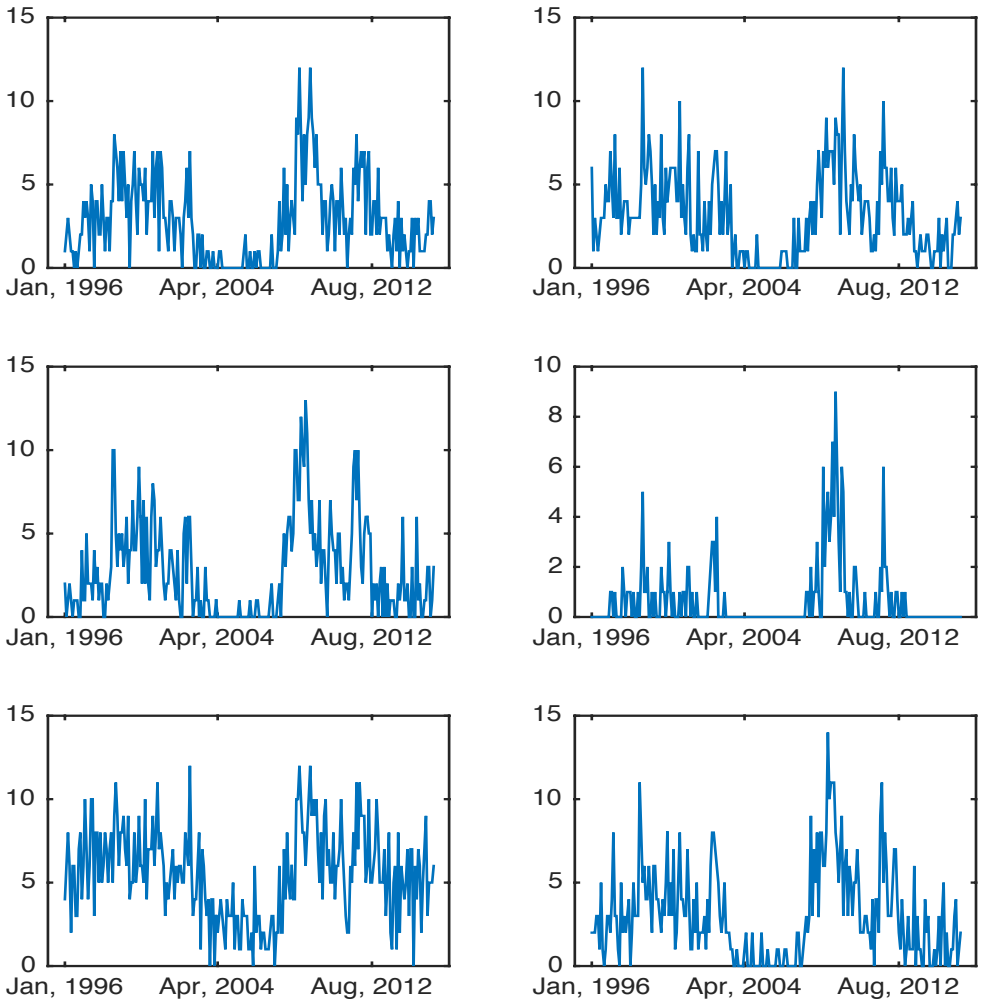

Figure 9.7: Exceedance number per month for BOA (left) and Citigroup (right) given different threshold (0.02, 0.02), (-0.02, 0.05), (0.01, -0.02) from top to bottom.

Table 9.3: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for BOA and Citigroup under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(-0.02, 0.01)			$\boldsymbol{\lambda}$=(0.02, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.361	0.794	6.50E-01	3.03E-07	0.073	1.00E+00
d_2	1.584	0.325	2.04E-06	2.29E-09	0.033	1.00E+00
a_{11}	0.562	0.219	1.10E-02	0.122	0.171	4.77E-01
a_{12}	-0.558	0.609	3.61E-01	0.467	0.175	8.08E-03
a_{21}	0.213	0.092	2.17E-02	-0.126	0.117	2.83E-01
a_{22}	-0.355	0.228	1.21E-01	0.580	0.127	7.69E-06
b_{11}	0.592	0.043	3.35E-32	0.291	0.038	7.22E-13
b_{12}	0.164	0.072	2.43E-02	0.047	0.042	2.65E-01
b_{21}	0.265	0.011	4.86E-65	0.182	0.027	1.76E-10
b_{22}	0.141	0.018	1.35E-13	0.301	0.039	3.34E-13
	$\boldsymbol{\lambda}$=(-0.02, 0.02)			$\boldsymbol{\lambda}$=(0.03, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	4.11E-05	0.081	1.00E+00	1.54E-04	0.566	1.00E+00
d_2	0.044	0.066	5.07E-01	0.530	0.657	4.20E-01
a_{11}	0.786	0.106	2.41E-12	0.912	0.394	2.15E-02
a_{12}	-0.587	0.148	9.24E-05	-0.601	0.682	3.79E-01
a_{21}	-0.046	0.082	5.73E-01	0.314	0.462	4.98E-01
a_{22}	0.424	0.112	2.02E-04	0.003	0.768	9.97E-01
b_{11}	0.453	0.030	4.25E-37	0.216	0.035	1.70E-09
b_{12}	0.270	0.040	9.54E-11	0.382	0.089	2.53E-05
b_{21}	0.292	0.021	1.06E-33	0.124	0.033	1.79E-04
b_{22}	0.249	0.029	2.41E-15	0.251	0.046	1.58E-07
	$\boldsymbol{\lambda}$=(-0.02, 0.05)			$\boldsymbol{\lambda}$=(0.01, -0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.453	0.099	6.96E-06	1.593	0.506	1.86E-03
d_2	0.005	0.295	9.87E-01	0.001	1.674	9.99E-01
a_{11}	-0.110	0.086	1.98E-01	-0.436	0.404	2.81E-01
a_{12}	0.088	0.058	1.32E-01	0.438	0.184	1.82E-02
a_{21}	-0.663	0.458	1.49E-01	-0.282	1.374	8.38E-01
a_{22}	0.569	0.243	1.99E-02	0.595	0.647	3.59E-01
b_{11}	0.540	0.041	2.77E-30	0.108	0.018	1.37E-08
b_{12}	0.256	0.030	3.09E-15	0.181	0.011	9.33E-39
b_{21}	0.191	0.242	4.31E-01	0.187	0.051	2.90E-04
b_{22}	0.716	0.247	4.18E-03	0.449	0.034	5.01E-30

Table 9.4: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for BOA and Citigroup under different threshold $\boldsymbol{\lambda}$ continued

	$\boldsymbol{\lambda}$=(0.01, 0.01)			$\boldsymbol{\lambda}$=(0.01, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.990	0.132	1.34E-36	0.916	0.123	1.47E-12
d_2	2.006	0.140	6.05E-34	0.916	0.136	1.23E-10
a_{11}	4.302	0.577	1.65E-12	-3.892	1.847	3.62E-02
a_{12}	-4.602	0.589	1.77E-13	4.116	1.784	2.19E-02
a_{21}	3.371	0.584	2.39E-08	-4.906	1.908	1.07E-02
a_{22}	-3.601	0.610	1.22E-08	5.087	1.849	6.41E-03
b_{11}	0.128	0.015	2.69E-16	0.044	0.011	1.69E-04
b_{12}	0.108	0.015	7.39E-12	0.182	0.013	1.63E-32
b_{21}	0.087	0.014	7.23E-10	0.053	0.012	9.22E-06
b_{22}	0.075	0.014	1.53E-07	0.213	0.014	1.81E-38
	$\boldsymbol{\lambda}$=(0.01, 0.02)			$\boldsymbol{\lambda}$=(0.02, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.236	0.308	7.91E-05	0.005	0.523	9.92E-01
d_2	0.154	0.405	7.05E-01	1.533	0.278	8.91E-08
a_{11}	-0.179	0.272	5.10E-01	0.775	0.187	4.55E-05
a_{12}	0.516	0.151	7.20E-04	-0.590	0.463	2.03E-01
a_{21}	-0.327	0.353	3.56E-01	0.465	0.079	1.30E-08
a_{22}	0.782	0.218	4.00E-04	-0.557	0.199	5.50E-03
b_{11}	0.065	0.021	1.83E-03	0.268	0.032	8.62E-15
b_{12}	0.086	0.015	2.58E-08	0.486	0.080	5.01E-09
b_{21}	0.116	0.035	9.95E-04	0.041	0.013	1.45E-03
b_{22}	0.332	0.040	1.38E-14	0.336	0.022	3.52E-38
	$\boldsymbol{\lambda}$=(0.01, 0.03)			$\boldsymbol{\lambda}$=(0.03, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.280	0.237	1.59E-07	0.046	0.750	9.51E-01
d_2	0.002	0.552	9.97E-01	0.773	0.326	1.85E-02
a_{11}	0.021	0.154	8.94E-01	0.843	0.137	3.06E-09
a_{12}	0.345	0.076	9.37E-06	-1.261	0.726	8.35E-02
a_{21}	-0.322	0.367	3.81E-01	0.125	0.072	8.66E-02
a_{22}	0.656	0.204	1.49E-03	0.198	0.197	3.18E-01
b_{11}	0.115	0.022	2.57E-07	0.196	0.088	2.73E-02
b_{12}	0.060	0.012	2.48E-06	1.053	0.509	3.95E-02
b_{21}	0.204	0.094	3.13E-02	-0.036	0.030	2.33E-01
b_{22}	0.399	0.067	7.28E-09	0.322	0.061	2.74E-07

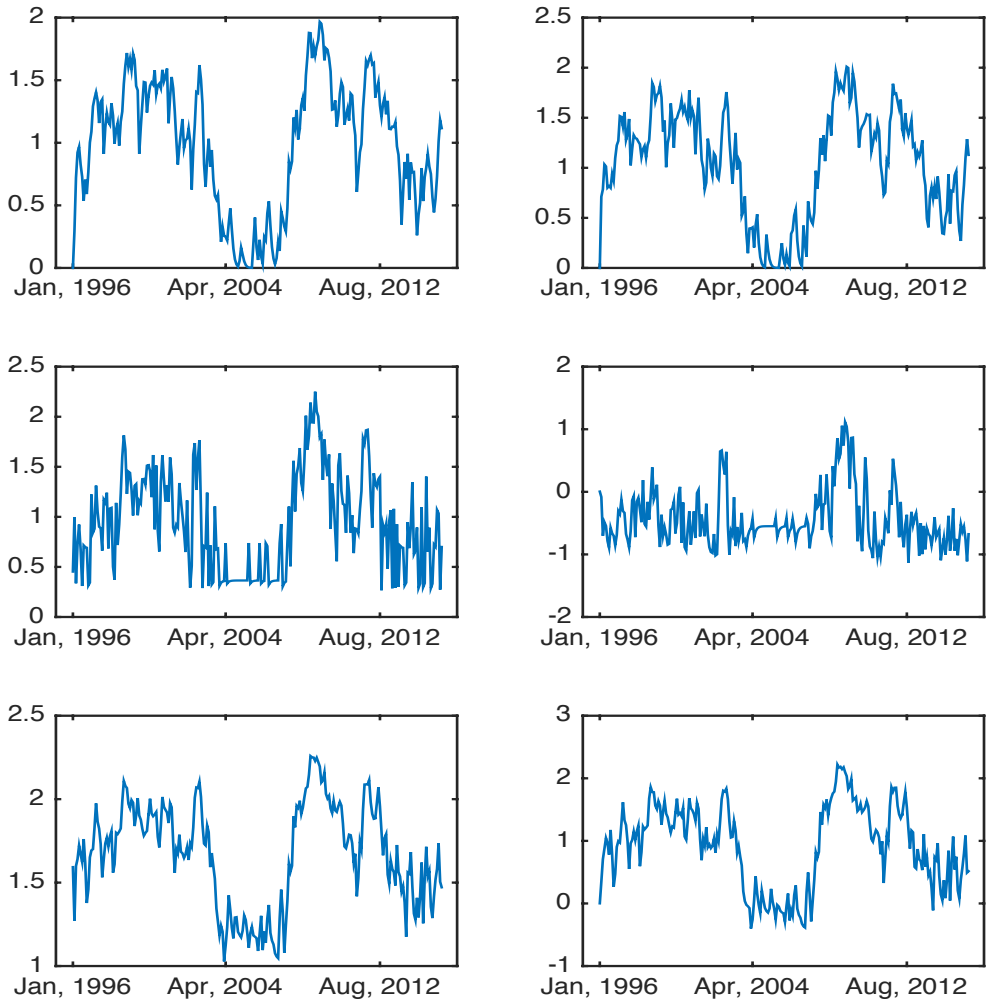

Figure 9.8: Time series of v_t in model 6.1.3 for BOA (left) and Citigroup (right) given different threshold (0.02, 0.02), (-0.02, 0.05), (0.01, -0.02) from top to bottom.

Figure 9.7 shows the plots for the exceedance number for BOA and Citigroup at different threshold, and the corresponding estimation results are shown in table 9.3 and table 9.4. Same as the analysis table for index return, the second column and the fifth column list the estimations from the maximum likelihood method. The third column and the sixth column show the standard error of the corresponding estimated coefficients. The fourth column and the last column in the table are the P-value of t-test statistic for the estimators. In table 9.3, under the paired thresholds $\boldsymbol{\lambda}$ = (-0.02, 0.01), from the estimated number of a_{12} and a_{21}, b_{12} and b_{21} in model 6.1.3, the interdependence between $v_{t+1,BOA}$ and $v_{t,Citi}$ is -0.558, the interdependence between $v_{t+1,BOA}$ and $v_{t,Citi}$ is 0.213, and the interdependence between $v_{t+1,BOA}$ and $log(Y_{t,Citi}+1)$ is 0.164, the interdependence between $v_{t+1,Citi}$ and $log(Y_{t,BOA}+1)$ is 0.265. Fixing the threshold for BOA at -0.02, change the threshold of Citigroup from 0.01 to 0.02, the interdependence between $v_{t+1,BOA}$ and $v_{t,Citi}$ becomes -0.587, the interdependence between $v_{t+1,Citi}$ and $v_{t,BOA}$ is -0.046, the interdependence between $v_{t+1,BOA}$ and $log(Y_{t,Citi}+1)$ becomes 0.270, and the interdependence between $v_{t+1,Citi}$ and $log(Y_{t,BOA}+1)$ becomes 0.292. While with changing both of the paired thresholds, say $\boldsymbol{\lambda}$ = (-0.02, 0.05) to $\boldsymbol{\lambda}$ = (0.01, -0.02), the parameters of a_{12}, a_{21}, b_{12}, which represent the interdependence in model 6.1.3, all have significant changes. This can also be seen from the middle and bottom plots of v_t in figure 9.8.

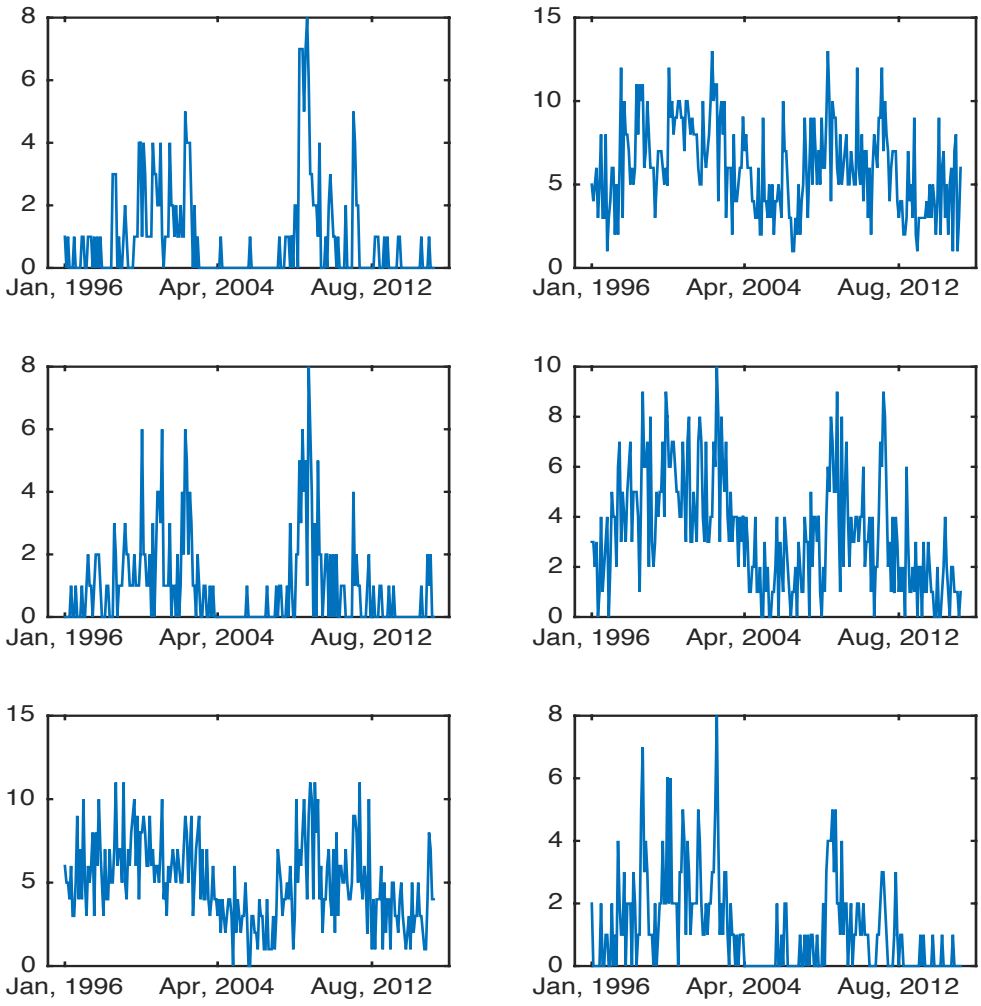

Figure 9.9: Exceedance number per month for GE (left) and Philips (right) given different threshold (-0.03, -0.01), (0.04, 0.02), (0.01, 0.04) from top to bottom.

Table 9.5: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for GE and Philips under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(-0.01, 0.01)			$\boldsymbol{\lambda}$=(0.01, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.752	0.422	7.59E-02	0.953	0.111	1.10E-15
d_2	1.373	0.673	4.24E-02	1.073	0.108	1.16E-19
a_{11}	1.172	0.249	4.18E-06	2.128	0.242	2.68E-16
a_{12}	-0.880	0.430	4.21E-02	-1.936	0.262	2.58E-12
a_{21}	0.639	0.380	9.37E-02	1.331	0.246	1.54E-07
a_{22}	-0.451	0.676	5.05E-01	-1.116	0.268	4.38E-05
b_{11}	0.231	0.012	5.20E-49	0.141	0.013	2.12E-23
b_{12}	0.119	0.018	1.58E-10	0.251	0.013	4.26E-50
b_{21}	0.119	0.010	2.76E-24	0.096	0.012	6.19E-14
b_{22}	0.074	0.016	8.32E-06	0.228	0.013	1.88E-46
	$\boldsymbol{\lambda}$=(0.01, 0.01)			$\boldsymbol{\lambda}$=(-0.03, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.423	0.411	6.42E-04	0.001	2.686	1.00E+00
d_2	1.732	0.334	4.77E-07	1.795	0.816	2.87E-02
a_{11}	2.201	0.470	4.68E-06	0.512	0.496	3.03E-01
a_{12}	-2.049	0.630	1.30E-03	-0.608	1.612	7.06E-01
a_{21}	1.563	0.402	1.31E-04	0.219	0.113	5.38E-02
a_{22}	-1.467	0.527	5.81E-03	-0.214	0.444	6.30E-01
b_{11}	0.159	0.013	1.61E-27	0.554	0.178	2.09E-03
b_{12}	0.077	0.015	1.12E-06	0.365	0.305	2.33E-01
b_{21}	0.096	0.011	4.85E-15	0.054	0.012	1.37E-05
b_{22}	0.071	0.015	1.79E-06	0.210	0.026	7.42E-14
	$\boldsymbol{\lambda}$=(0.01, -0.02)			$\boldsymbol{\lambda}$=(-0.01, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.438	1.006	1.54E-01	2.167	0.072	7.74E-83
d_2	0.913	1.737	6.00E-01	1.715	0.109	5.88E-39
a_{11}	-0.644	1.043	5.37E-01	1.112	0.053	1.31E-55
a_{12}	0.663	0.595	2.66E-01	-1.842	0.077	2.44E-65
a_{21}	-1.275	1.759	4.69E-01	0.262	0.042	2.28E-09
a_{22}	1.267	0.985	1.99E-01	-0.407	0.094	2.36E-05
b_{11}	0.118	0.021	6.76E-08	0.352	0.019	2.43E-47
b_{12}	0.193	0.014	1.41E-31	0.352	0.022	7.24E-17
b_{21}	0.245	0.038	4.72E-10	0.137	0.011	2.43E-29
b_{22}	0.310	0.038	4.01E-23	0.091	0.013	1.11E-10

Table 9.6: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for GE and Philips under different threshold $\boldsymbol{\lambda}$ continued

	$\boldsymbol{\lambda}$=(0.01, -0.03)			$\boldsymbol{\lambda}$=(0.02, 0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.710	0.617	6.06E-03	0.075	0.651	9.08E-01
d_2	0.314	1.655	8.50E-01	1.838	0.651	8.18E-10
a_{11}	-0.354	0.438	4.19E-01	0.521	0.183	4.77E-03
a_{12}	0.325	0.167	5.24E-02	-0.027	0.431	9.50E-01
a_{21}	-0.969	1.049	3.56E-01	0.356	0.067	2.29E-07
a_{22}	0.726	0.440	1.00E-01	-0.219	0.180	2.25E-01
b_{11}	0.109	0.022	1.95E-06	0.540	0.066	2.27E-14
b_{12}	0.153	0.017	4.96E-17	-0.143	0.081	7.80E-02
b_{21}	0.582	0.159	3.13E-04	0.072	0.009	1.01E-12
b_{22}	0.423	0.046	2.75E-17	0.031	0.018	8.69E-02
	$\boldsymbol{\lambda}$=(0.01, 0.02)			$\boldsymbol{\lambda}$=(0.03, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.608	1.465	2.73E-01	0.002	0.644	9.98E-01
d_2	1.095	2.489	6.61E-01	1.003	0.406	1.42E-02
a_{11}	-0.981	1.660	5.55E-01	0.782	0.319	1.49E-02
a_{12}	1.067	1.008	2.91E-01	-0.816	0.467	8.16E-02
a_{21}	-1.557	2.767	5.74E-01	0.392	0.176	2.70E-02
a_{22}	1.691	1.705	3.22E-01	-0.093	0.304	7.61E-01
b_{11}	0.143	0.021	8.40E-11	0.365	0.097	2.16E-04
b_{12}	0.049	0.018	6.01E-03	0.547	0.214	1.12E-02
b_{21}	0.270	0.038	1.72E-11	0.075	0.029	1.00E-02
b_{22}	0.089	0.018	1.64E-06	0.235	0.044	2.84E-07
	$\boldsymbol{\lambda}$=(0.01, 0.04)			$\boldsymbol{\lambda}$=(-0.02, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.191	0.290	5.38E-05	0.316	0.657	6.32E-01
d_2	4.42E-05	1.405	1.00E+00	2.396	0.162	1.92E-35
a_{11}	0.142	0.187	4.49E-01	0.683	0.133	6.03E-07
a_{12}	0.200	0.099	4.38E-02	-0.681	0.433	1.18E-01
a_{21}	-0.581	0.748	4.39E-01	0.373	0.036	1.32E-20
a_{22}	0.838	0.318	8.95E-03	-0.708	0.089	6.39E-14
b_{11}	0.059	0.024	1.68E-02	0.409	0.047	5.88E-16
b_{12}	0.125	0.018	4.41E-11	0.371	0.095	1.17E-04
b_{21}	0.475	0.232	4.22E-02	0.070	0.015	2.85E-06
b_{22}	0.205	0.067	2.52E-03	0.214	0.025	4.97E-16

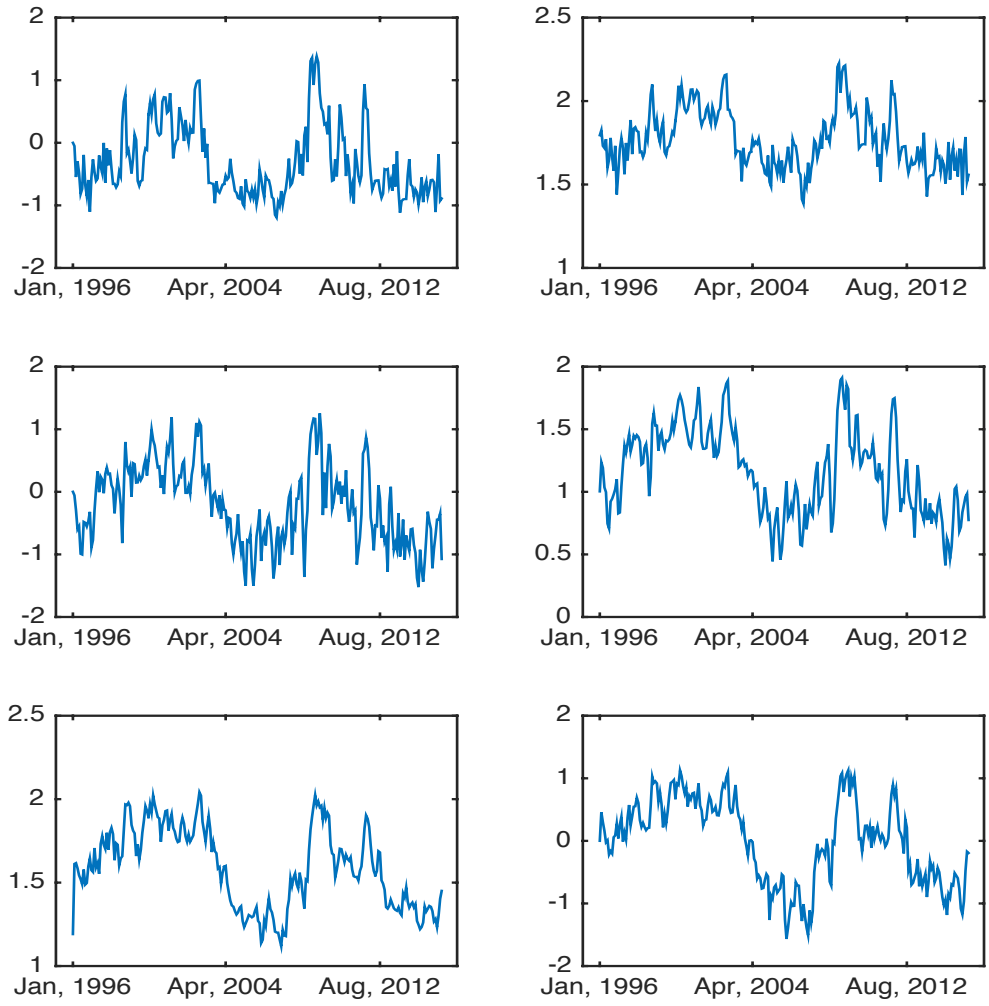

Figure 9.10: Time series of v_t in model 6.1.3 for GE (left) and Philips (right) given different threshold (-0.03, -0.01), (0.03, 0.02), (0.01, 0.04) from top to bottom.

We also choose GE and Philips in machine industry to do the same analysis. The plots of the number of exceedance return for GE and Philips under different threshold are shown in figure 9.9. Table 9.5 and table 9.6 are the estimation results given different threshold sets, figure 9.10 plots the time series of $v_{t,GE}$ and $v_{t,Philips}$. Similar as the results for commercial bank industry, many of the estimations of the cross term are significant, which means there do have an interdependence for the exceedance return between the two stocks. In addition, the exceedance event can be seen as a tail problem in stock return, which indirectly shows an interdependence for the tail stock return problem.

Similar results are shown from the petroleum industry in the following with observation plots given different threshold sets in figure 9.11, and the estimation results are given in table 9.7, table 9.8. The plots for $v_{t,Marathon}$ and $v_{t,Exxon}$ is shown in figure 9.12.

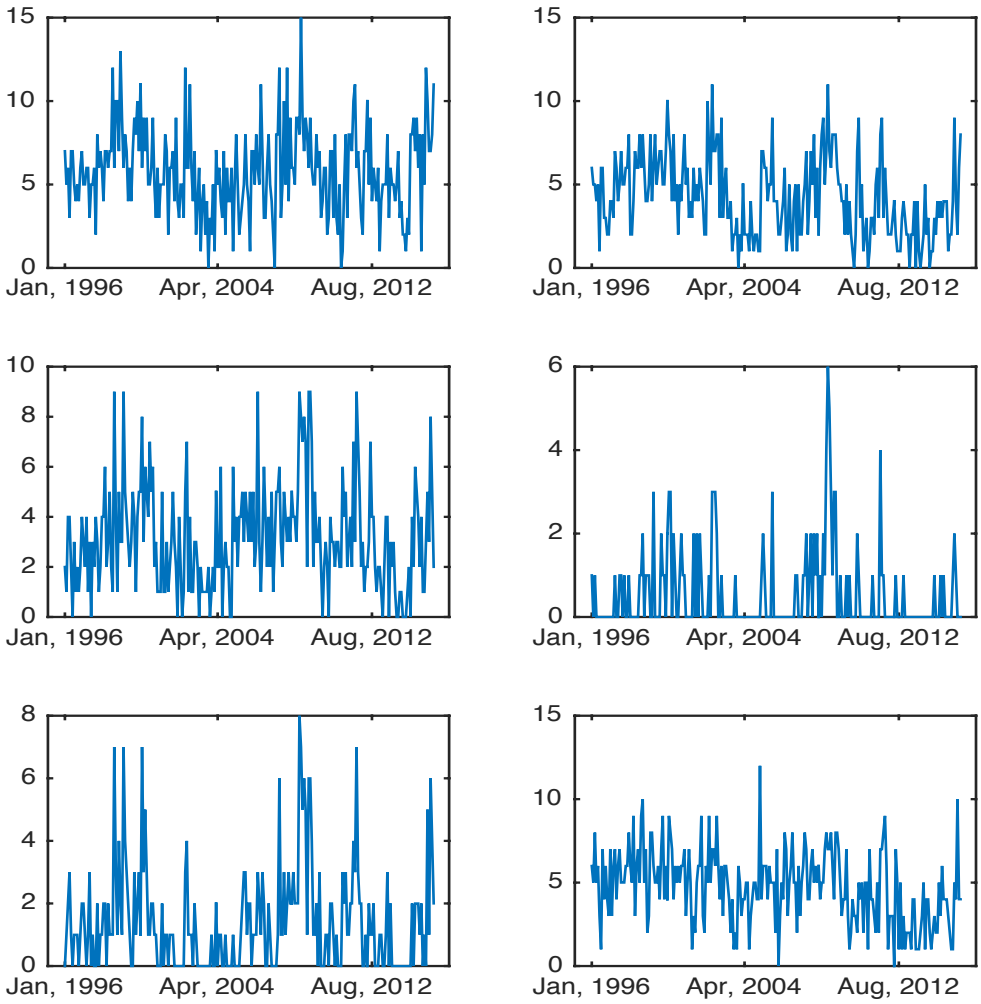

Figure 9.11: Exceedance number per month for Marathon Oil (left) and Exxon Mobil (right) given different threshold (-0.01, -0.01), (0.02, -0.03), (0.03, 0.01) from top to bottom.

Table 9.7: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for Marathon Oil and Exxon Mobil under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(-0.01, -0.01)			$\boldsymbol{\lambda}$=(0.02, -0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.991	0.128	2.57E-13	0.405	0.214	5.95E-02
d_2	0.621	0.176	4.84E-04	0.474	0.300	1.15E-01
a_{11}	0.097	0.111	3.86E-01	-0.581	0.984	5.56E-01
a_{12}	0.097	0.053	1.85E-04	0.689	0.662	2.99E-01
a_{21}	-0.549	0.158	6.13E-04	-1.848	1.346	1.71E-01
a_{22}	0.787	0.074	7.16E-22	1.699	0.903	6.10E-02
b_{11}	0.111	0.013	1.24E-14	0.092	0.014	1.55E-10
b_{12}	0.075	0.012	4.55E-09	0.205	0.019	7.32E-22
b_{21}	0.108	0.019	1.86E-08	0.132	0.016	6.38E-15
b_{22}	0.290	0.020	1.25E-33	0.316	0.020	4.80E-38
	$\boldsymbol{\lambda}$=(-0.02, -0.01)			$\boldsymbol{\lambda}$=(0.02, -0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.041	0.072	5.64E-01	0.048	0.046	3.05E-01
d_2	0.405	0.062	4.70E-10	5.44E-09	0.079	1.00E+00
a_{11}	0.662	0.065	1.41E-20	0.698	0.057	2.23E-27
a_{12}	-0.241	0.111	3.16E-02	-0.204	0.043	3.40E-06
a_{21}	0.184	0.044	4.38E-05	-0.260	0.097	7.94E-03
a_{22}	0.152	0.069	2.96E-02	0.597	0.073	1.64E-14
b_{11}	0.332	0.030	7.92E-23	0.144	0.027	2.95E-07
b_{12}	0.160	0.042	1.95E-04	0.247	0.021	1.02E-25
b_{21}	0.041	0.017	1.30E-02	0.091	0.056	1.04E-01
b_{22}	0.368	0.025	6.45E-35	0.437	0.054	2.31E-14
	$\boldsymbol{\lambda}$=(0.03, -0.01)			$\boldsymbol{\lambda}$=(0.02, -0.03)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.022	0.552	9.68E-01	0.514	0.239	3.20E-02
d_2	0.984	0.266	2.64E-04	2.65E-04	0.782	1.00E+00
a_{11}	0.797	0.209	1.70E-04	0.358	0.151	1.84E-02
a_{12}	-0.749	0.506	1.40E-01	0.083	0.098	3.95E-01
a_{21}	0.293	0.082	4.46E-04	-0.885	0.703	2.09E-01
a_{22}	-0.122	0.195	5.30E-01	0.681	0.324	3.66E-02
b_{11}	0.265	0.074	4.18E-04	0.158	0.059	7.78E-03
b_{12}	0.578	0.157	2.95E-04	0.219	0.050	1.76E-05
b_{21}	-0.019	0.018	2.85E-01	0.442	0.430	3.05E-01
b_{22}	0.373	0.030	1.93E-27	0.596	0.232	1.09E-02

Table 9.8: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for Marathon Oil and Exxon Mobil under different threshold $\boldsymbol{\lambda}$ continued

	$\boldsymbol{\lambda}$=(-0.03, -0.01)			$\boldsymbol{\lambda}$=(-0.01, 0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.004	0.307	9.91E-01	0.588	0.040	1.76E-35
d_2	0.595	0.139	2.86E-05	0.572	0.040	2.29E-33
a_{11}	0.694	0.138	9.51E-07	0.309	0.042	2.37E-12
a_{12}	-0.523	0.319	1.02E-01	0.014	0.030	6.33E-01
a_{21}	0.144	0.054	7.47E-03	-0.665	0.041	7.98E-41
a_{22}	0.168	0.113	1.40E-01	0.979	0.031	2.83E-88
b_{11}	0.368	0.059	1.99E-09	0.213	0.011	9.15E-51
b_{12}	0.366	0.120	2.53E-03	0.129	0.011	1.03E-26
b_{21}	0.054	0.016	1.15E-03	0.196	0.011	5.90E-44
b_{22}	0.330	0.028	6.31E-26	0.162	0.012	4.41E-31
	$\boldsymbol{\lambda}$=(-0.03, -0.02)			$\boldsymbol{\lambda}$=(0.03, 0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.003	0.192	9.89E-01	0.007	0.288	9.82E-01
d_2	0.147	0.177	4.07E-01	0.441	0.069	7.40E-10
a_{11}	1.053	0.449	1.98E-02	0.567	0.091	1.89E-09
a_{12}	-0.910	0.675	1.79E-01	-0.316	0.241	1.91E-01
a_{21}	0.394	0.386	3.08E-01	-0.009	0.020	6.38E-01
a_{22}	-0.266	0.507	6.00E-01	0.425	0.055	3.28E-13
b_{11}	0.412	0.063	3.13E-10	0.457	0.083	9.08E-08
b_{12}	0.202	0.074	6.62E-03	0.163	0.128	2.04E-01
b_{21}	0.275	0.055	9.79E-07	0.061	0.014	1.01E-05
b_{22}	0.248	0.059	3.48E-05	0.245	0.021	1.68E-24
	$\boldsymbol{\lambda}$=(-0.04, -0.02)			$\boldsymbol{\lambda}$=(0.02, 0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	3.67E-06	0.812	1.00E+00	1.254	0.511	1.49E-02
d_2	0.313	0.485	5.19E-01	1.805	0.303	9.62E-09
a_{11}	1.037	0.673	1.25E-01	1.807	0.437	4.88E-05
a_{12}	-1.264	0.830	1.29E-01	-1.676	0.641	9.48E-03
a_{21}	0.293	0.414	4.79E-01	1.303	0.322	7.13E-05
a_{22}	-0.120	0.577	8.35E-01	-1.318	0.406	1.33E-03
b_{11}	0.608	0.132	6.96E-06	0.181	0.021	1.44E-15
b_{12}	0.446	0.217	4.13E-02	0.118	0.019	4.98E-09
b_{21}	0.389	0.082	3.10E-06	0.167	0.015	3.68E-24
b_{22}	0.255	0.092	5.97E-03	0.061	0.019	1.19E-03

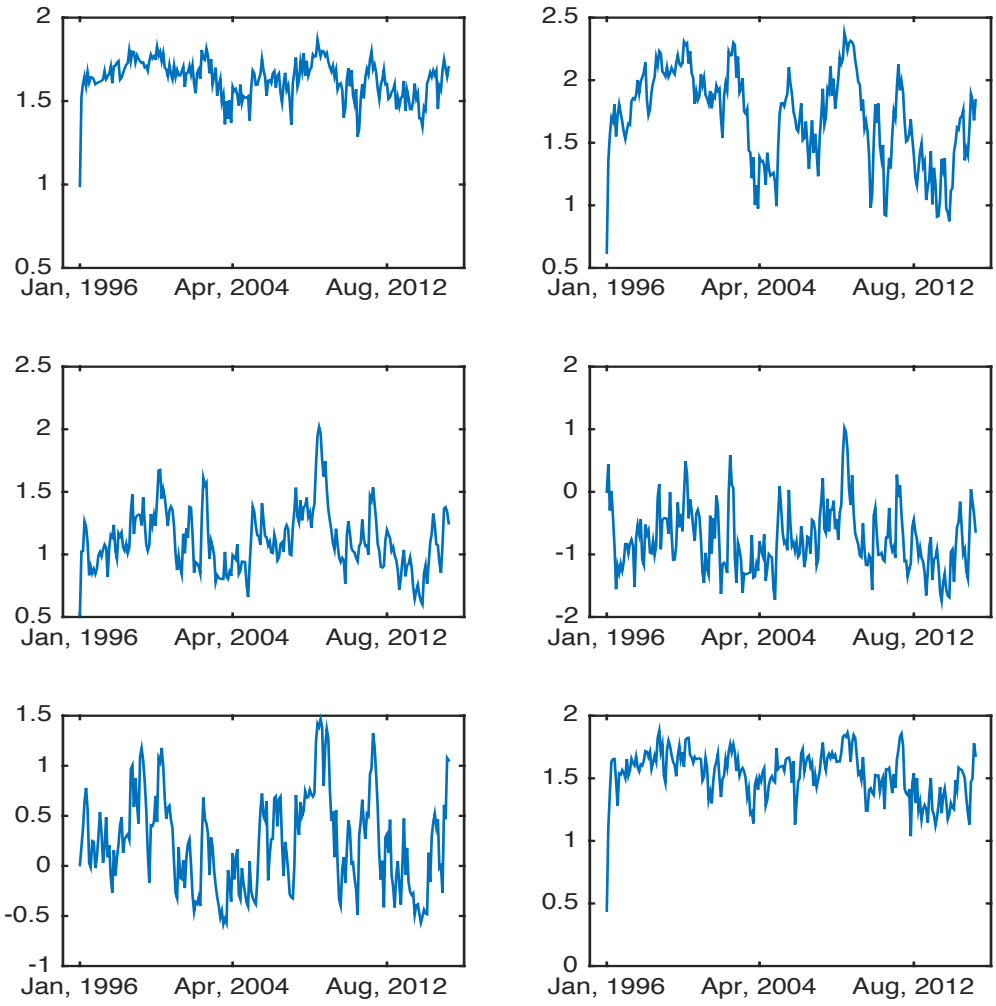

Figure 9.12: Time series of v_t in model 6.1.3 for Marathon Oil (left) and Exxon Mobil (right) given different threshold (-0.01, -0.01), (0.02, -0.03), (0.03, 0.01) from top to bottom.

9.2.2 Application to the dependence financial risks in different market

Not only for the dependence of different index/stock in the same market/industry, an interesting application of this model is to analyze the dependence of financial risks in different market. We choose SP500 index in the US, FTSE 100 in London stock exchange, and Nikkei 225 in Japan market in this study to analyze the interdependence of this three markets under different risk levels (i.e., different thresholds). Figure 9.13 is the 20 years' daily index return for the three indices respectively.

We first apply the bivariate log-Poisson model to SP500 and FTSE 100, SP500 and Nikkei 225 respectively under different threshold. Figure 9.14 and 9.16 list the exceedance number per month for SP500 (left) and FTSE 100 (right), for SP500 (left) and Nikkei 225 (right) under different threshold, respectively. And table 9.9 and 9.10 are the results for the bivariate model applications. In addition, figure 9.15 and 9.17 are the corresponding time series of v_t for the two examples.

Then we apply the three dimensional log-Poisson model to this three indices together under different threshold to analyze the dependence in the US market, the England market and the Japan market. Table 9.11 and table 9.12 are the estimation results for the parameters in model 8.2.1 under different threshold. We also plot v_t of SP500, FTSE 100 and Nikkei 225 from this regression which list in figure 9.18.

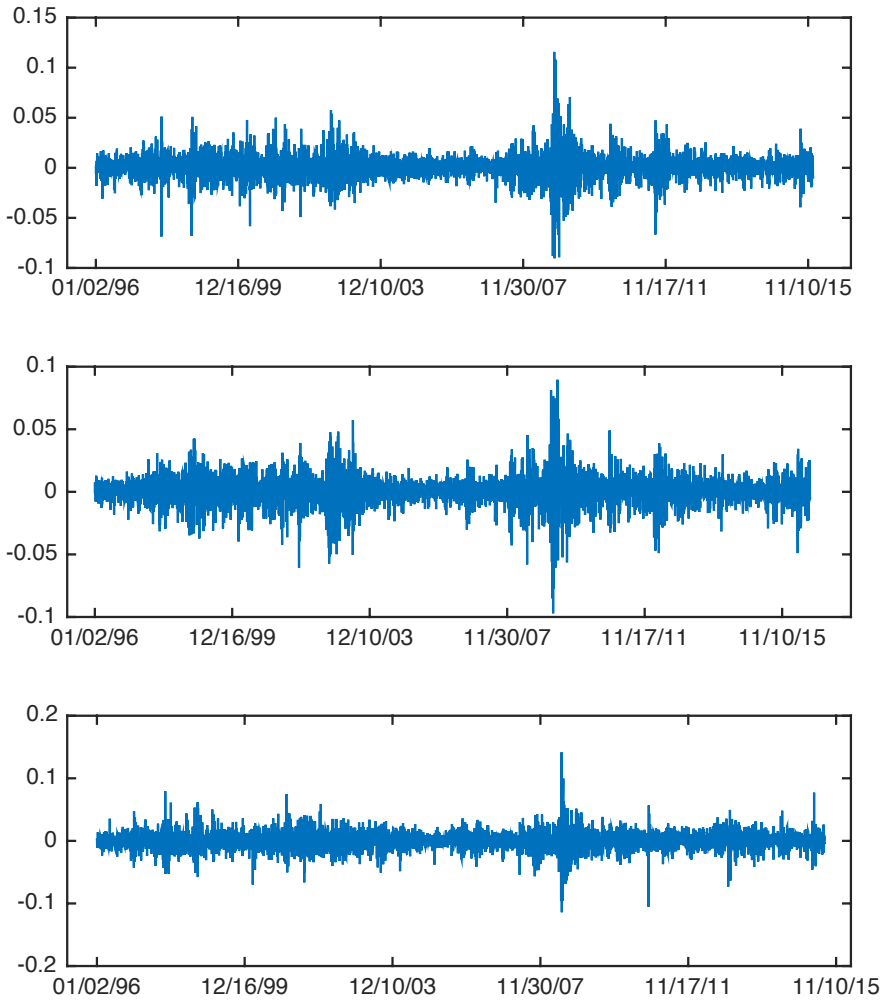

Figure 9.13: 20 Years daily return for SP500, FTSE 100 and Nikkei 225 (from top to bottom), the horizontal axis is the time period from Jan 1996 to Dec 2015, the vertical axis is the daily return.

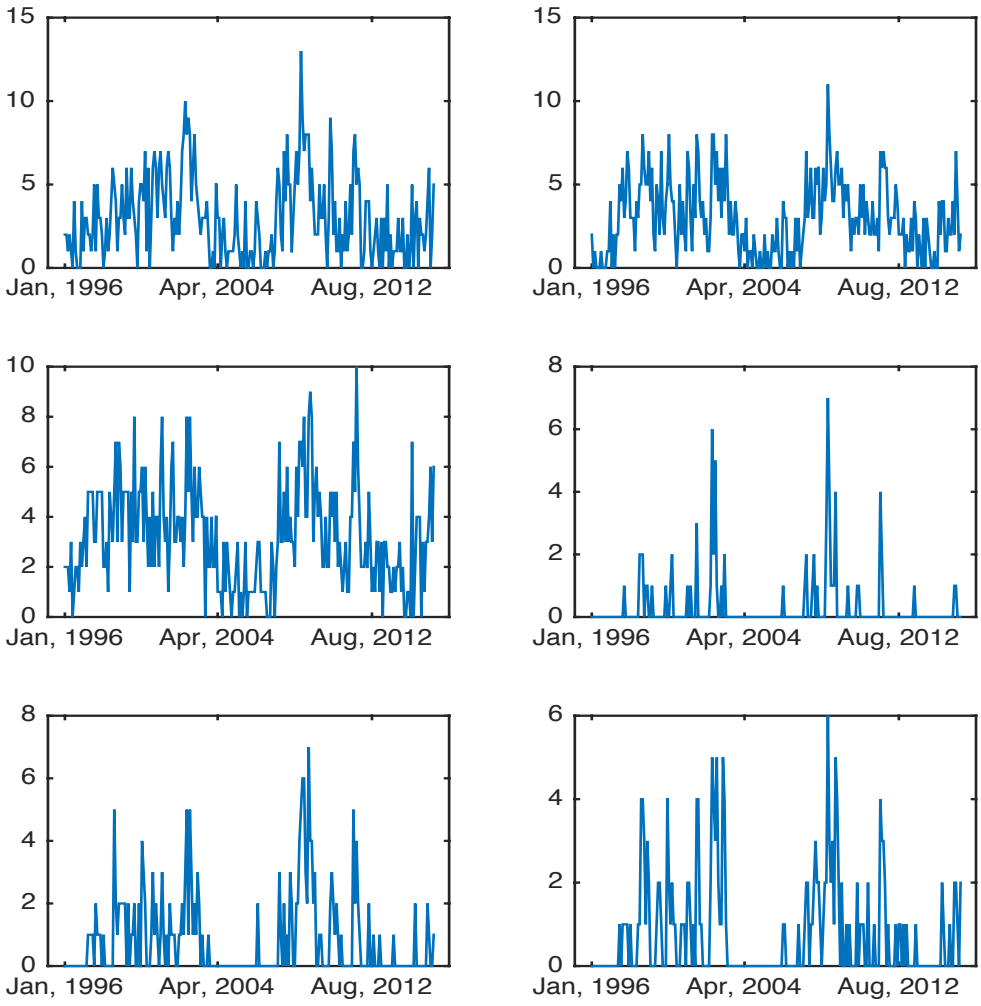

Figure 9.14: Exceedance number per month for SP500 (left) and FTSE 100 (right) given different threshold (-0.01, 0.01), (0.01, -0.03), (0.02, 0.02) from top to bottom.

Table 9.9: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for SP500 and FTSE 100 under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(0.01, 0.01)			$\boldsymbol{\lambda}$=(-0.01, 0.01)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.636	0.282	2.49E-02	0.092	0.069	1.81E-01
d_2	0.260	0.321	4.18E-01	3.92E-08	0.050	1.00E+00
a_{11}	-1.368	0.815	9.46E-02	-0.031	0.184	8.69E-01
a_{12}	1.535	0.662	2.13E-02	0.459	0.170	7.64E-03
a_{21}	-1.363	0.976	1.64E-01	-0.673	0.189	4.40E-04
a_{22}	1.597	0.795	4.56E-02	1.127	0.160	1.94E-11
b_{11}	0.055	0.038	1.53E-01	0.432	0.031	1.75E-32
b_{12}	0.303	0.033	1.99E-17	-0.006	0.030	8.36E-01
b_{21}	0.204	0.048	3.52E-05	0.369	0.030	8.45E-28
b_{22}	0.312	0.036	6.24E-16	0.102	0.028	2.82E-04
	$\boldsymbol{\lambda}$=(0.01, -0.02)			$\boldsymbol{\lambda}$=(-0.02, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	2.086	0.721	4.16E-03	0.178	2.266	9.37E-01
d_2	0.916	1.267	4.70E-01	3.14E-04	3.658	1.00E+00
a_{11}	-0.896	0.586	1.27E-01	-1.415	6.056	8.15E-01
a_{12}	0.756	0.430	7.95E-02	0.941	3.628	7.96E-01
a_{21}	-1.533	0.831	6.61E-02	-3.147	10.123	7.56E-01
a_{22}	1.091	0.600	7.02E-02	2.064	5.956	7.29E-01
b_{11}	0.004	0.046	9.33E-01	0.539	0.122	1.49E-05
b_{12}	0.389	0.031	1.59E-27	0.257	0.147	8.05E-02
b_{21}	0.397	0.261	1.30E-01	0.953	0.265	3.99E-04
b_{22}	0.597	0.094	9.13E-10	0.336	0.274	2.21E-01
	$\boldsymbol{\lambda}$=(0.01, -0.03)			$\boldsymbol{\lambda}$=(0.02, 0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.471	0.330	1.54E-01	0.148	1.293	9.09E-01
d_2	4.63E-04	1.327	1.00E+00	6.98E-06	2.822	1.00E+00
a_{11}	0.326	0.167	5.23E-02	-1.332	4.770	7.80E-01
a_{12}	0.020	0.114	8.63E-01	0.728	2.211	7.42E-01
a_{21}	-1.269	1.429	3.76E-01	-3.983	10.137	6.95E-01
a_{22}	0.631	0.423	1.37E-01	2.057	4.735	6.64E-01
b_{11}	0.201	0.116	8.35E-02	0.170	0.154	2.71E-01
b_{12}	0.326	0.084	1.31E-04	0.417	0.144	4.20E-03
b_{21}	0.620	1.182	6.00E-01	0.515	0.268	5.61E-02
b_{22}	0.842	0.435	5.39E-02	0.780	0.257	2.72E-03

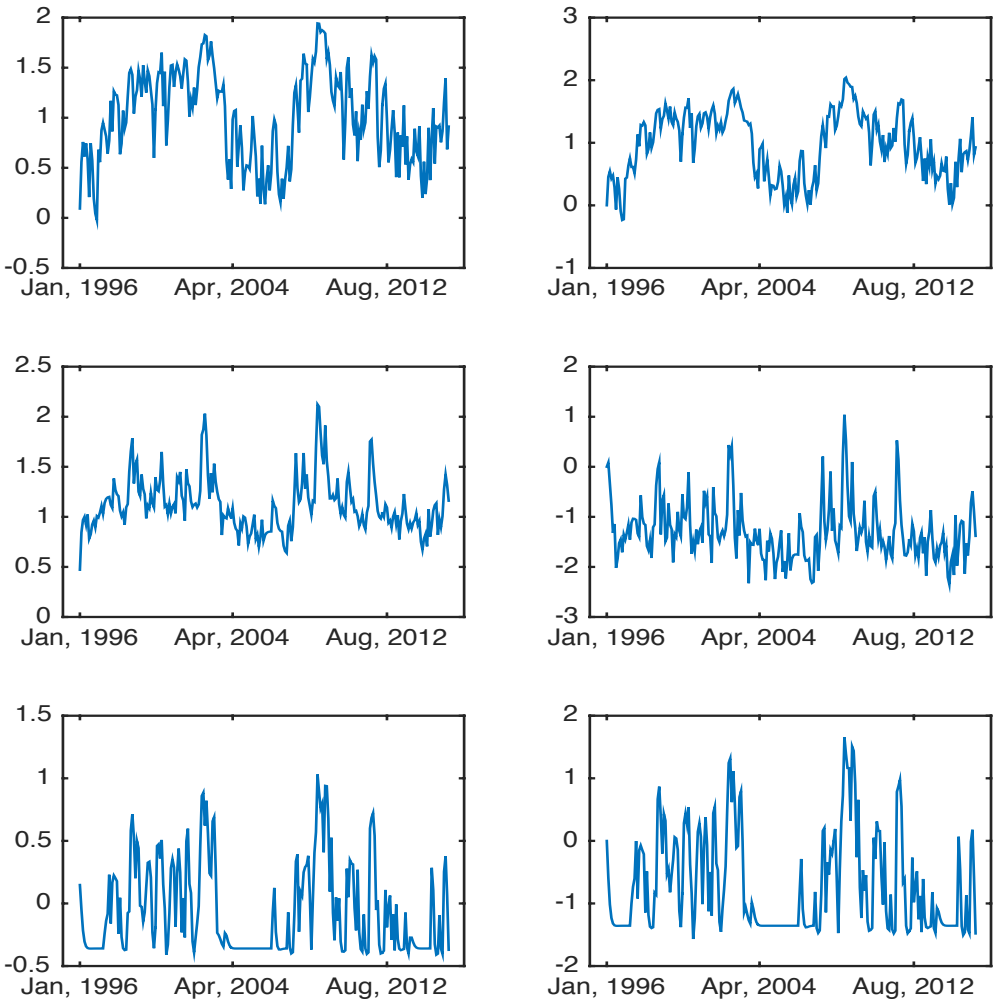

Figure 9.15: Time series of v_t in model 6.1.3 for SP500 (left) and FTSE 100 (right) given different threshold (-0.01, 0.01), (0.01, -0.03), (0.02, 0.02) from top to bottom.

From the results table in 9.9, the off-diagonal entries a_{12}, a_{21} of the parameter matrix $\boldsymbol{A}$ and b_{12}, b_{21} of the parameter matrix $\boldsymbol{B}$ which we care more in this model have significant estimations. This shows the index return tail problem in the US market are interacted by the exceedance return of the European market. Specifically, with the threshold for FTSE 100 fixed at 0.01, the exceedance return for the index SP500 has different estimations of the cross terms under this model with threshold at 0.01 to -0.01. The interdependence between $v_{t+1,SP500}$ and $v_{t,FTSE100}$ representing by a_{12} decrease from 1.535 to 0.459, the interdependence between $v_{t+1,FTSE100}$ and $v_{t,SP500}$ representing by a_{21} changes from -1.363 to -0.673 , the interdependence between $v_{t+1,SP500}$ and $log(Y_{t,FTSE100}+1)$ representing by b_{12} changes from positive 0.303 to negative of -0.006, the interdependence between $v_{t+1,FTSE100}$ and $log(Y_{t,SP500}+1)$ increases from 0.204 to 0.369 with very significant estimation. More results within different paired thresholds are given in the same table.

For the interdependance between US market and the Japan Market, the same estimation under this bivariate log-liner Poisson autoregression are analyzed in table 9.10. The results show an interaction of the exceedance index return between the US market and Japan market.

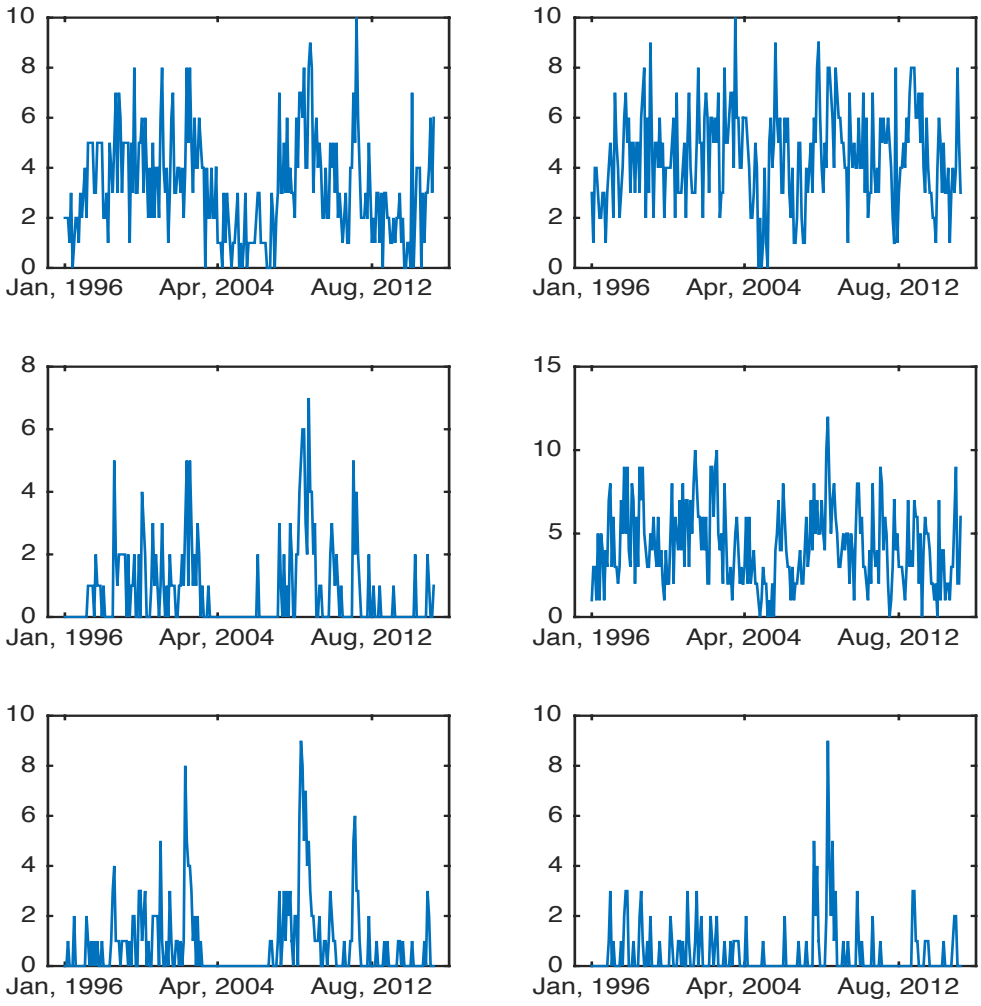

Figure 9.16: Exceedance number per month for SP500 (left) and Nikkei 225 (right) given different threshold (0.01, 0.01), (0.02, -0.01), (-0.02, -0.03) from top to bottom.

Table 9.10: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for SP500 and Nikkei 225 under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(0.01, 0.01)			$\boldsymbol{\lambda}$=(0.01, -0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.137	0.066	3.84E-02	1.92E-04	0.035	9.96E-01
d_2	0.320	0.052	2.57E-09	1.57E-10	0.058	1.00E+00
a_{11}	0.705	0.035	2.99E-54	0.618	0.046	1.15E-30
a_{12}	-0.127	0.081	1.20E-01	-0.147	0.044	8.92E-04
a_{21}	0.006	0.022	8.00E-01	-0.125	0.092	1.75E-01
a_{22}	0.474	0.052	3.14E-17	0.587	0.085	3.97E-11
b_{11}	0.322	0.026	4.37E-27	0.269	0.031	2.79E-16
b_{12}	-0.025	0.036	4.95E-01	0.186	0.020	1.31E-17
b_{21}	0.082	0.015	1.76E-07	0.047	0.063	4.53E-01
b_{22}	0.214	0.022	3.35E-19	0.350	0.053	2.13E-10
	$\boldsymbol{\lambda}$=(0.02, 0.01)			$\boldsymbol{\lambda}$=(0.02, -0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.405	0.986	6.82E-01	3.03E-04	0.424	9.99E-01
d_2	0.716	0.109	2.76E-10	0.296	0.220	1.80E-01
a_{11}	0.207	0.187	2.71E-01	0.635	0.200	1.69E-03
a_{12}	-0.447	0.688	5.16E-01	-1.088	0.737	1.41E-01
a_{21}	-0.010	0.023	6.54E-01	-0.079	0.116	4.95E-01
a_{22}	0.221	0.084	8.79E-03	-0.021	0.315	9.47E-01
b_{11}	0.898	0.180	1.24E-06	0.523	0.128	6.39E-05
b_{12}	-0.204	0.282	4.70E-01	0.406	0.151	7.62E-03
b_{21}	0.095	0.016	2.67E-08	0.264	0.062	2.99E-05
b_{22}	0.255	0.027	9.76E-19	0.220	0.058	1.78E-04
	$\boldsymbol{\lambda}$=(0.02, -0.01)			$\boldsymbol{\lambda}$=(-0.02, -0.03)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.195	1.015	8.48E-01	8.53E-07	0.096	1.00E+00
d_2	0.789	0.288	6.63E-03	2.43E-10	0.106	1.00E+00
a_{11}	0.444	0.276	1.09E-01	-0.048	0.189	7.98E-01
a_{12}	-0.681	0.716	3.43E-01	-0.695	0.332	3.75E-02
a_{21}	0.064	0.078	4.17E-01	-0.289	0.194	1.38E-01
a_{22}	0.224	0.192	2.46E-01	-0.230	0.250	3.59E-01
b_{11}	0.732	0.164	1.20E-05	0.455	0.085	1.91E-07
b_{12}	0.236	0.222	2.88E-01	0.156	0.079	4.91E-02
b_{21}	0.166	0.019	4.54E-16	-0.179	0.130	1.70E-01
b_{22}	0.195	0.028	3.51E-11	0.416	0.141	3.61E-03

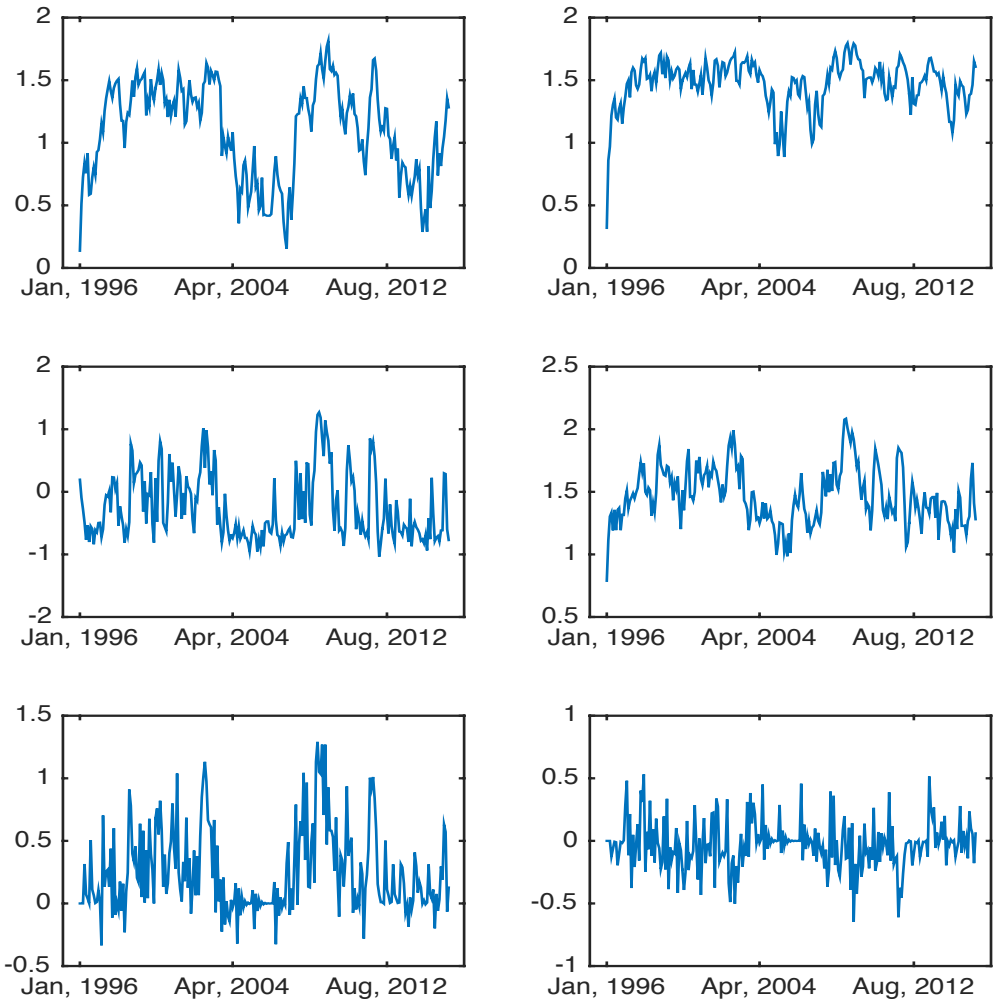

Figure 9.17: Time series of v_t in model 6.1.3 for SP500 (left) and Nikkei 225 (right) given different threshold (0.01, 0.01), (0.02, -0.01), (-0.02, -0.03) from top to bottom.

To clearly analyze the interdependence of financial risks from the exceedance index return among the US market, the European market and the Japan market, we apply the three-dimensional log-linear Poisson autoregression model to estimate the cross terms for the three indices simultaneously. Table 9.11 shows the estimation results under the threshold sets of (0.01, 0.01, 0.01) and (0.01, 0.02, 0.03) for SP500, FTSE100 and Nikkei 225 respectively. With the threshold fixed at 0.01 for SP500 and changing the threshold of FTSE100 from 0.01 to 0.02, Nikkei from 0.01 to 0.03, the impact to $v_{t+1,SP500}$ from $v_{t,FTSE100}$ has decreased from -0.982 to -0.147, and the impact from $v_{t,Nikkei225}$ changes from -0.587 to -0.147. From the estimation of a_{21}, a_{23}, the the impact to $v_{t+1,FTSE100}$ from $v_{t,SP500}$ has a big change from 3.475 to -0.125, and the impact from $v_{t,Nikkei225}$ changes from -1.856 to 0.587. The interaction to FTSE100 from the US market and Japan market are stronger than the interaction to SP500 from FTSE100 and Nikkei225. Similar analysis and results can be seen from the estimation of the entries of the parameter matrix $\boldsymbol{B}$ in the table and some of the estimations for the cross terms are significant which can be seen from the P-value of t-statistic. More estimation results under different threshold sets are given in table 9.12. The time series of $v_{t,SP500}$, $v_{t,FTSE100}$ and $v_{t,Nikkei225}$ are plotted in figure 9.18.

Table 9.11: Estimation Results of Three Dimentional Log-linear Poisson Autoregression Model for SP500, FTSE 100 and Nikkei 225 under different threshold $\boldsymbol{\lambda}$

	$\boldsymbol{\lambda}$=(0.01, 0.01, 0.01)			$\boldsymbol{\lambda}$=(0.01, 0.02, 0.03)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.379	0.326	2.46E-01	0.714	0.627	2.56E-01
d_2	1.198	0.702	8.91E-02	0.012	1.042	9.91E-01
d_3	0.202	0.116	8.30E-02	0.002	2.462	9.99E-01
a_{11}	1.897	0.646	3.66E-03	0.086	0.452	8.49E-01
a_{12}	-0.982	0.415	1.89E-02	-0.147	0.292	4.93E-01
a_{13}	-0.587	0.425	1.68E-01	-0.147	0.148	1.88E-01
a_{21}	3.475	1.168	3.25E-03	-0.125	0.715	3.15E-01
a_{22}	-1.740	0.738	1.92E-02	0.587	0.509	2.10E-01
a_{23}	-1.856	0.806	2.22E-02	0.587	0.327	8.76E-01
a_{31}	-0.330	0.228	1.49E-01	-0.125	1.818	7.01E-01
a_{32}	0.190	0.146	1.94E-01	0.587	1.000	7.79E-01
a_{33}	0.727	0.147	1.42E-06	0.587	0.539	9.44E-01
b_{11}	0.156	0.033	4.57E-06	0.269	0.062	5.41E-01
b_{12}	0.283	0.025	1.37E-23	0.186	0.050	1.07E-15
b_{13}	-0.037	0.035	2.90E-01	0.186	0.039	8.09E-02
b_{21}	0.076	0.043	7.41E-02	0.047	0.340	3.55E-01
b_{22}	0.355	0.041	1.09E-15	0.350	0.177	4.16E-04
b_{23}	-0.031	0.047	5.11E-01	0.350	0.210	9.29E-01
b_{31}	0.008	0.019	6.77E-01	0.047	0.470	7.51E-01
b_{32}	0.113	0.016	1.19E-11	0.350	0.347	3.24E-01
b_{33}	0.141	0.021	2.20E-10	0.350	0.409	7.51E-02

Table 9.12: Estimation Results of Three Dimentional Log-linear Poisson Autoregression Model for SP500, FTSE 100 and Nikkei 225 under different threshold $\boldsymbol{\lambda}$ (continued)

	$\boldsymbol{\lambda}$=(-0.01, 0.01, -0.02)			$\boldsymbol{\lambda}$=(0.01, -0.01, -0.02)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.135	0.139	3.32E-01	0.141	0.265	5.96E-01
d_2	0.095	0.130	4.65E-01	0.196	0.339	5.64E-01
d_3	1.03E-08	0.093	1.00E+00	1.77E-08	0.257	1.00E+00
a_{11}	-0.838	0.688	2.24E-01	2.939	1.285	2.31E-02
a_{12}	1.567	0.739	3.49E-02	-2.545	1.340	5.87E-02
a_{13}	-0.638	0.278	2.23E-02	-0.138	0.229	5.46E-01
a_{21}	-1.666	0.717	2.10E-02	2.851	1.227	2.10E-02
a_{22}	2.304	0.762	2.79E-03	-2.387	1.288	6.50E-02
a_{23}	-0.555	0.293	6.00E-02	-0.247	0.281	3.79E-01
a_{31}	-1.130	0.453	1.34E-02	2.065	1.235	9.59E-02
a_{32}	0.830	0.474	8.15E-02	-2.278	1.267	7.35E-02
a_{33}	0.443	0.185	1.74E-02	0.541	0.233	2.10E-02
b_{11}	0.360	0.029	5.44E-28	0.025	0.033	4.45E-01
b_{12}	-0.017	0.030	5.80E-01	0.410	0.029	6.38E-33
b_{13}	0.097	0.024	5.07E-05	0.050	0.024	3.70E-02
b_{21}	0.369	0.030	7.56E-28	-0.047	0.034	1.67E-01
b_{22}	0.042	0.031	1.75E-01	0.398	0.030	9.05E-30
b_{23}	0.102	0.024	3.12E-05	0.087	0.024	3.67E-04
b_{31}	0.266	0.051	3.32E-07	-0.081	0.081	3.16E-01
b_{32}	0.104	0.045	2.25E-02	0.255	0.056	7.96E-06
b_{33}	0.181	0.043	3.48E-05	0.284	0.063	8.98E-06

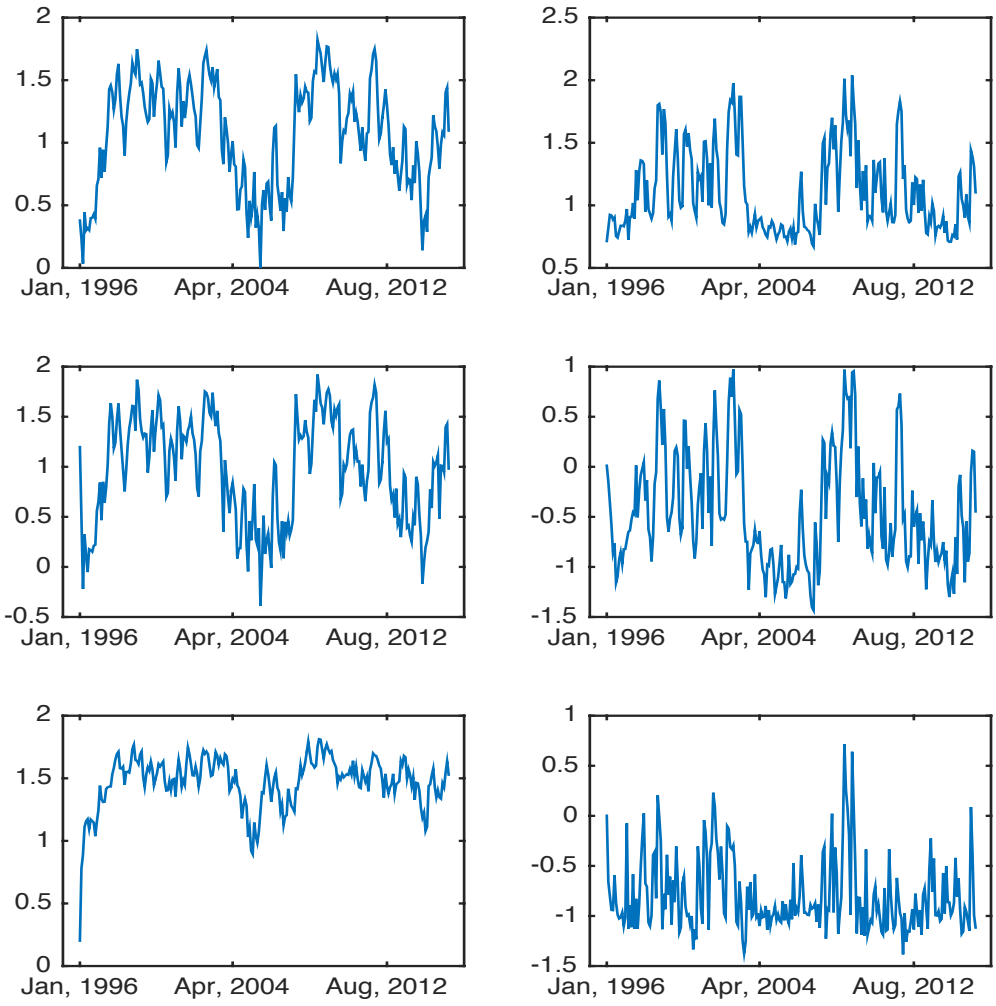

Figure 9.18: Time series of v_t in model 8.2.1 for SP500, FTSE 100 and Nikkei 225 (from top to bottom) given different threshold (0.01, 0.01, 0.01) (left), (0.01, 0.02, 0.03) (right).

9.3 Trading Transactions Examples

Another application of this multivariate linear log-Poisson autoregression is the interdependence of trading behavior for different stocks, through which the impact of one stock's trading behavior on another stock can be quantitatively modeled and identified by this model. TAQ data of the high frequency transactions per minute or within two minutes or five minutes are used as the observations in this study. The transactions in Jan 2002 and Jan 2012 of Bank of America vs Citigroup, GE vs Philips are considered in the analysis of this model. As in section 9.2, to clearly analyze the result and the interdependence of between the transactions of different stocks, we first plot the transactions number per minute in figure 9.19 and 9.21, and list the estimation results for the parameters in table 9.13 and table 9.14, then plot the v_t of the model for the two industries respectively. In Jan 2002, the data set are considered of 21 transaction days of total 8190 observations for one-minute data, 4095 observations for two-minute data, and 1638 observations for five-minute data, while in Jan 2012, 20 transaction days of total 7800 observations for one-minute, 3900 observations for two-minute, and 1560 observations for five-minute are considered in this study.

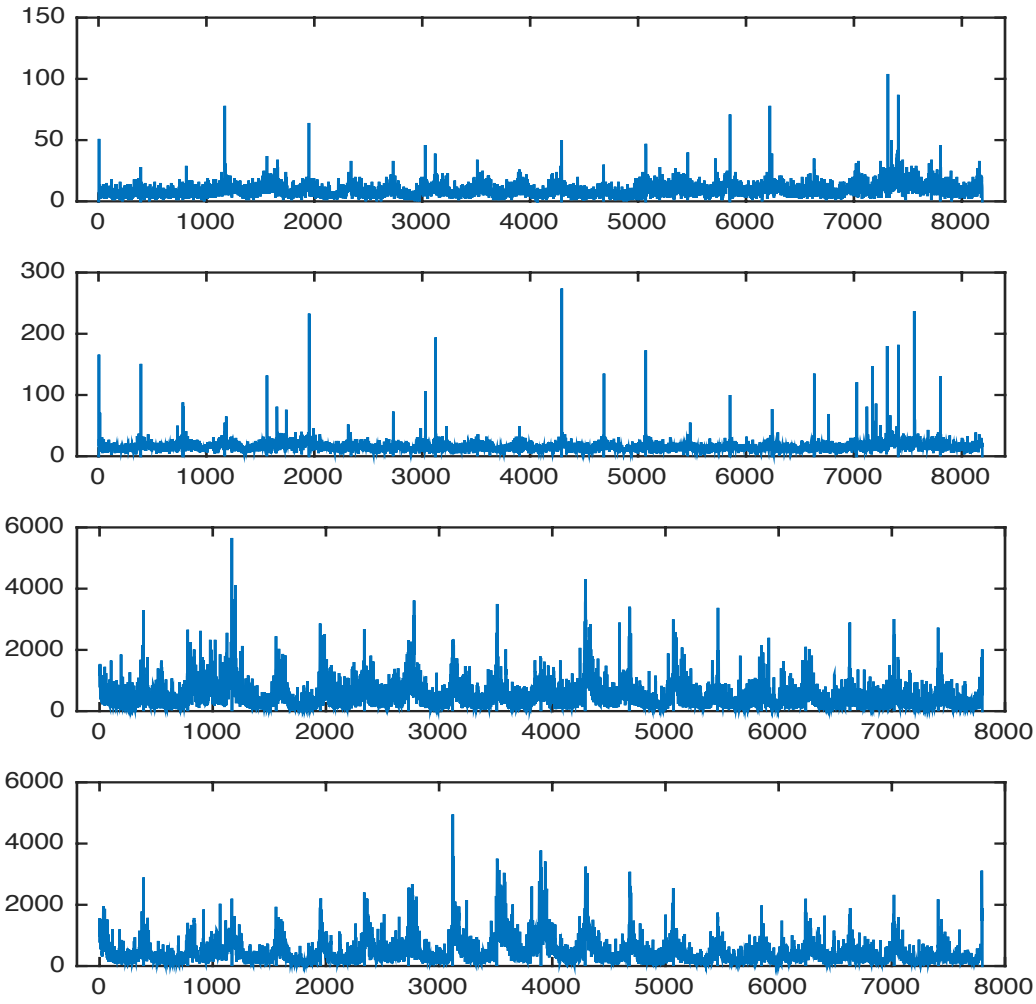

Figure 9.19: Transactions per minute for BOA and Citigroup in Jan 2002 (top two), and in Jan 2012 (bottom two). The horizontal axis is time, the vertical axis is the number of transactions.

Table 9.13: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for Transactions for BOA and Citigroup

	1-minute (Jan,2002)			1-minute (Jan,2012)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	1.2958	0.0353	7.87E-273	0.2202	5.46E-06	0.0000
d_2	2.1868	0.0208	0.0000	0.2002	5.89E-06	0.0000
a_{11}	1.2765	0.0095	0.0000	0.8511	2.07E-06	0.0000
a_{12}	-0.7897	0.0208	1.19E-290	-0.0401	1.85E-06	0.0000
a_{21}	0.5200	0.0064	0.0000	-0.0521	2.16E-06	0.0000
a_{22}	-0.2986	0.0123	2.61E-126	0.8197	2.13E-06	0.0000
b_{11}	0.0891	0.0008	0.0000	0.1176	1.39E-06	0.0000
b_{12}	0.0339	0.0003	0.0000	0.0412	1.49E-06	0.0000
b_{21}	0.0969	0.0005	0.0000	0.0502	1.42E-06	0.0000
b_{22}	0.0233	0.0004	0.0000	0.1552	1.74E-06	0.0000
	2-minute (Jan,2002)			2-minute (Jan,2012)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	4.2103	0.0192	0.0000	0.2984	8.64E-06	0.0000
d_2	4.5351	0.0129	0.0000	0.2778	5.66E-06	0.0000
a_{11}	2.3861	0.0077	0.0000	0.8012	2.39E-06	0.0000
a_{12}	-2.3498	0.0116	0.0000	0.0148	1.33E-06	0.0000
a_{21}	1.5297	0.0058	0.0000	-0.0471	1.66E-06	0.0000
a_{22}	-1.5112	0.0083	0.0000	0.8004	1.37E-06	0.0000
b_{11}	0.0244	0.0003	0.0000	0.1430	1.34E-06	0.0000
b_{12}	-0.0478	0.0003	0.0000	0.0018	1.05E-06	0.0000
b_{21}	-0.0069	0.0002	3.12E-155	0.0378	9.99E-07	0.0000
b_{22}	-0.0746	0.0002	0.0000	0.1729	1.09E-06	0.0000
	5-minute (Jan,2002)			5-minute (Jan,2012)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	3.7640	0.0113	0.0000	5.4439	1.43E-05	0.0000
d_2	4.2923	0.0104	0.0000	5.3208	1.21E-05	0.0000
a_{11}	1.5018	0.0019	0.0000	3.6515	1.68E-05	0.0000
a_{12}	-1.3941	0.0042	0.0000	-3.4083	1.73E-05	0.0000
a_{21}	0.6676	0.0018	0.0000	2.6655	1.66E-05	0.0000
a_{22}	-0.5970	0.0039	0.0000	-2.3805	1.71E-05	0.0000
b_{11}	0.0746	0.0001	0.0000	-0.0931	3.89E-07	0.0000
b_{12}	0.0290	0.0001	0.0000	0.1243	3.91E-07	0.0000
b_{21}	0.0104	0.0001	0.0000	-0.1039	3.84E-07	0.0000
b_{22}	0.0211	0.0001	0.0000	0.1086	3.89E-07	0.0000

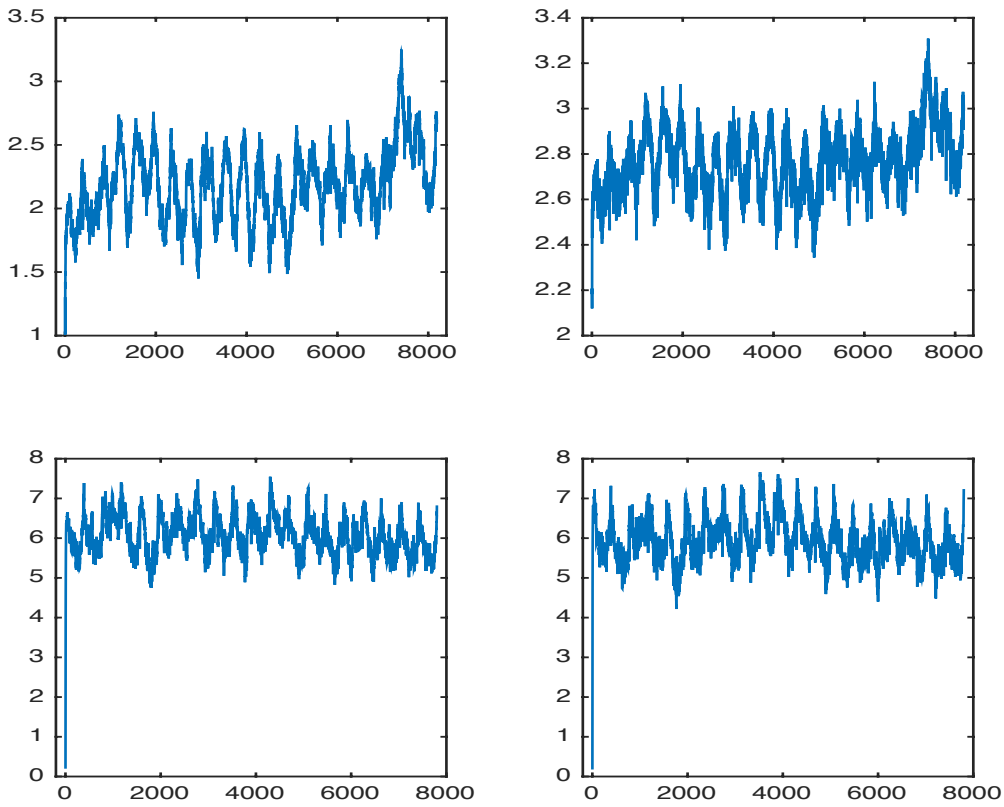

Figure 9.20: Time series of v_t in model 6.1.3 for BOA (left) and Citigroup (right) of one-minute transactions in Jan 2002 and Jan 2012 (from top to bottom).

In the result table 9.13 for BOA and Citigroup, from the estimated number of a_{12} and a_{21}, b_{12} and b_{21} in model 6.1.3, in Jan 2002, for the one-minute time interval transaction data, the interdependence between $v_{t+1,BOA}$ and $v_{t,Citi}$ is -0.7897, the

interdependence between $v_{t+1,Citi}$ and $v_{t,BOA}$ is 0.5200, the interdependence between $v_{t+1,BOA}$ and $log(Y_{t,Citi}+1)$ is 0.0339, and the interdependence between $v_{t+1,Citi}$ and $log(Y_{t,BOA}+1)$ is 0.0969; while, in Jan 2012, the correlation between $v_{t+1,BOA}$ and $v_{t,Citi}$ becomes less dependent at -0.0401, the interdependence between $v_{t+1,Citi}$ and $v_{t,BOA}$ becomes negative to -0.0521, and the interdependence between $v_{t+1,BOA}$ and $log(Y_{t,Citi}+1)$ changes to 0.0412, the correlation between $v_{t+1,Citi}$ and $log(Y_{t,BOA}+1)$ becomes less dependent to 0.0502. Comparing this results to the two-minutes and five-minutes table, the impact to $v_{t+1,BOA}$ from $v_{t,Citi}$ and the impact to $v_{t+1,Citi}$ from $v_{t,BOA}$ are both getting stronger with the value of the observations getting larger. The time series of $v_{t+1,BOA}$ and $v_{t+1,Citi}$ in Jan 2002 and Jan 2012 for one-minute transactions are plotted in figure 9.20.

Except for the commercial banking industry, more regression results in machine industry are given in table 9.14 and figure 9.21 and 9.22.

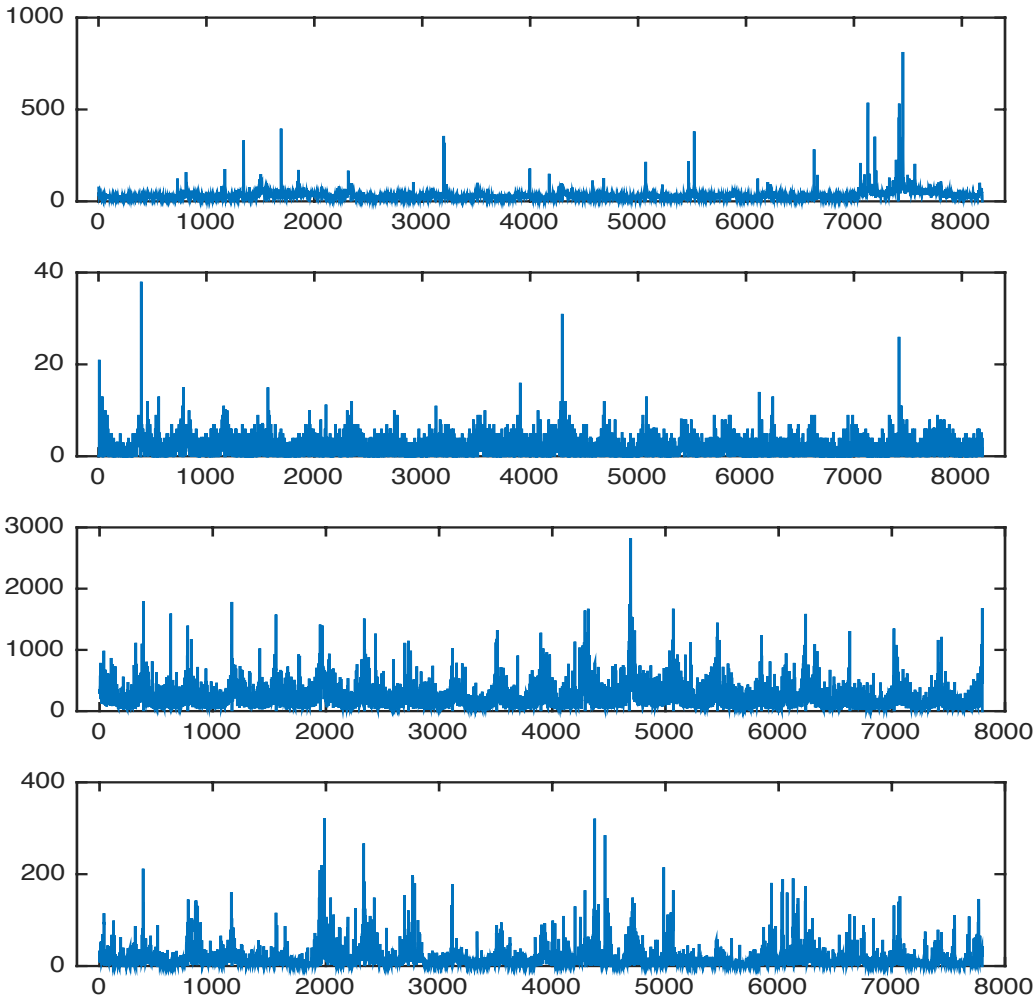

Figure 9.21: Transactions per minute for GE and Philips in Jan 2002 (top two), and in Jan 2012 (bottom two). The horizontal axis is time, the vertical axis is the number of transactions.

Table 9.14: Estimation Results of Bivariate Log-linear Poisson Autoregression Model for Transactions for GE and Philips

	1-minute (Jan,2002)			1-minute (Jan,2012)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.0233	0.0012	1.07E-87	4.3549	0.0119	0.0000
d_2	4.94E-11	0.0050	1.0000	0.0143	0.0169	0.3983
a_{11}	0.7782	0.0005	0.0000	-0.3863	0.0036	0.0000
a_{12}	0.0287	0.0005	0.0000	0.8985	0.0029	0.0000
a_{21}	-0.0546	0.0057	1.29E-21	-0.1181	0.0050	2.99E-118
a_{22}	0.8810	0.0037	0.0000	0.9427	0.0037	0.0000
b_{11}	0.2176	0.0003	0.0000	0.1435	1.34E-05	0.0000
b_{12}	-0.0241	0.0002	0.0000	0.0079	7.09E-06	0.0000
b_{21}	0.0450	0.0056	8.33E-16	0.1381	0.0001	0.0000
b_{22}	0.1230	0.0036	6.06E-241	0.0265	0.0001	0.0000
	2-minute (Jan,2002)			2-minute (Jan,2012)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	0.0427	0.0018	1.48E-114	4.0540	0.0082	0.0000
d_2	1.91E-12	0.0039	1.00E+00	0.0138	0.0112	0.2157
a_{11}	0.7582	0.0006	0.0000	-0.1572	0.0022	0.0000
a_{12}	0.0782	0.0008	0.0000	0.6594	0.0015	0.0000
a_{21}	-0.0052	0.0023	2.32E-02	-0.1711	0.0028	0.0000
a_{22}	0.8366	0.0021	0.0000	0.9182	0.0020	0.0000
b_{11}	0.2234	0.0003	0.0000	0.1369	5.73E-06	0.0000
b_{12}	-0.0449	0.0001	0.0000	0.0012	6.68E-06	0.0000
b_{21}	0.0149	0.0022	1.81E-11	0.1960	4.86E-05	0.0000
b_{22}	0.1282	0.0019	0.0000	0.0464	0.0001	0.0000
	5-minute (Jan,2002)			5-minute (Jan,2012)		
	MLE	Standard Error	P-value	MLE	Standard Error	P-value
d_1	3.1772	0.0492	0.0000	6.7165	0.0298	0.0000
d_2	5.3022	0.0468	0.0000	0.0055	0.1281	9.65E-01
a_{11}	0.5304	0.0055	0.0000	-0.3151	0.0055	0.0000
a_{12}	-0.6621	0.0099	0.0000	0.4315	0.0021	0.0000
a_{21}	-0.4759	0.0048	0.0000	-0.1348	0.0241	2.60E-08
a_{22}	-0.1356	0.0111	0.0000	0.8055	0.0097	0.0000
b_{11}	0.1278	0.0001	0.0000	0.0753	7.45E-06	0.0000
b_{12}	0.0232	0.0001	0.0000	0.0452	6.51E-06	0.0000
b_{21}	-0.0589	0.0005	0.0000	0.1710	0.0002	0.0000
b_{22}	-0.0219	0.0003	1.05E-32	0.1484	0.0002	0.0000

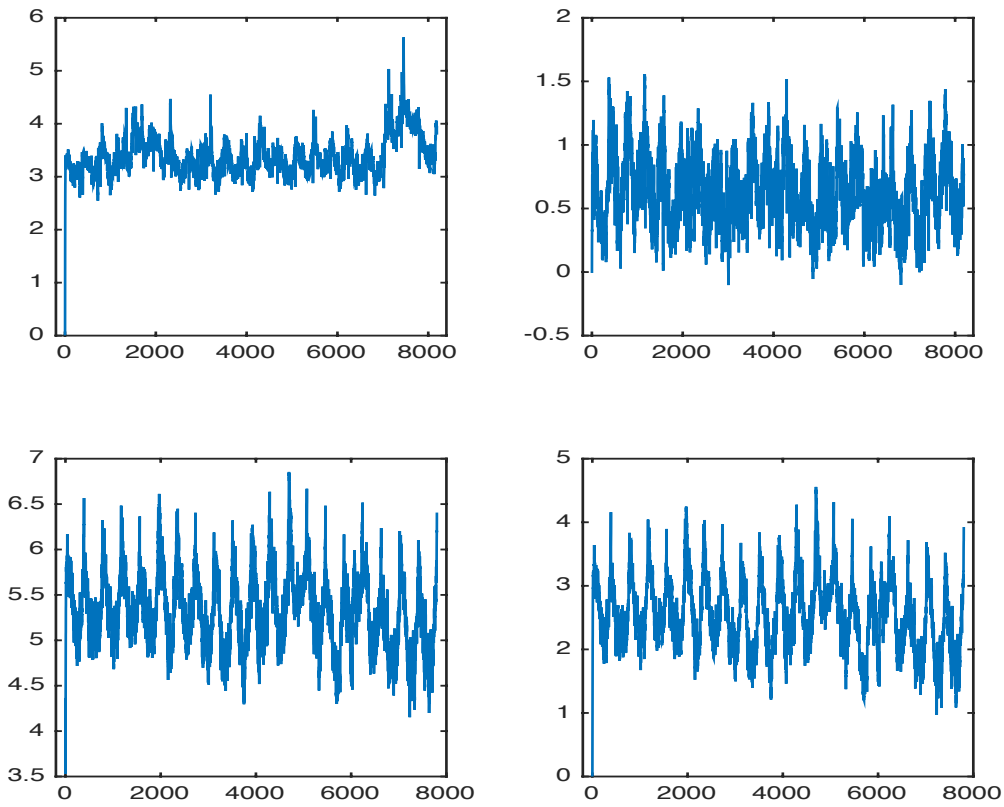

Figure 9.22: Time series of v_t in model 6.1.3 for GE (left) and Philips (right) of one-minute transactions in Jan 2002 and Jan 2012 (from top to bottom).

In table 9.14, the interdependence of the transactions between GE and Philips is clearly described from the estimation of the parameters in model 6.1.3, where a_{11} and a_{12} describe the interdependence of v_{t+1} and v_t between GE and Philips, while b_{12}

and b_{21} describe the interdependence of v_{t+1} and $log(Y_t+1)$ between GE and Philips. Similar as previous results, the coefficients of the interdependence have a significant change from 2002 to 2012, and a small change with the time interval changing. In this example, since the observations of Philips are much fewer than the observations of GE which cause an insignificant estimation of $d2$, which is also shown from the value of vertical-axis in figure 9.21.

9.4 Concluding Remarks

This multivariate log-liner Poisson autoregression model has solved the drawback of only positive association taken into account in Poisson autoregression. Moreover, comparing with the univariate log-liner Poisson autoregression model, this multivariate model has been applied to the application of the dependence of different stocks in the same industry, especially the dependence of financial risks in different market. In addition, it is applied to the interdependence of trading behavior for different stocks, through which to study the impact of one stock's trading behavior on another stock. Massive real data examples in difference cases and industries have been quantitatively modeled in this study and the interdependence of these problems are clearly seen from the estimation results under this model.

Chapter 10

Appendix

In this appendix, some theorems and definitions which are used in previous chapters are given from references including Meyn and Tweedie (2005), Wu and Shao (2004), Diaconis and Freedman (1999) and Doukhan and Wintenberger (2008). The proofs can be found in the original references and will not be restated here.

First, we list some definitions and theorems described in Meyn and Tweedie (2005), which we have referred to and used in the previous chapter.

Definition 10.0.1. State Space Definitions
(i) The state space X is called *countable* if X is discrete, with a finite or countable number of elements, and with $\mathcal{B}(X)$ the σ-field of all subsets of X.

(ii) The state space X is called *general* if it is equipped with a countably generated σ-field $\mathcal{B}(X)$.

(iii) The state space X is called *topological* if it is equipped with a locally compact, separatable, metrizable topology with $\mathcal{B}(X)$ as the Borel σ-field.

Definition 10.0.2. If $P = \{P(x, A), x \in X, A \in \mathcal{B}(X)\}$ is such that

(i) for each $A \in \mathcal{B}(X)$, $P(\cdot, A)$ is a non-negative measurable function on X

(ii) for each $x \in X$, $P(x, \cdot)$ is a probability measure on $\mathcal{B}(X)$

then we call P a transition probability kernel or Markov transition function.

Definition 10.0.3. A Markov chain $\{X_t\}$ with state space X is *ϕ-irreducible* if there exists a measure ϕ on $\mathcal{B}(X)$ such that

$$\sum_{t=1}^{\infty} P^t(x, A) > 0, \ forall x \in X$$

whenever $\phi(A) > 0$, where $A \in \mathcal{B}(X)$.

Definition 10.0.4. If $P(\cdot, O)$ is a lower semicontinuous function for any open set $O \in \mathcal{B}(X)$, then P is called a (weak) Feller chain.

The transition probability kernel P acts on bounded functions through the mapping is

$$Ph(x) = \int P(x, dy)h(y), x \in X$$

Suppose that X is a topological space, and we denote the class of bounded continuous from X to $\mathbb{R}$ by $\mathcal{C}$.

If the transition probability kernel P maps from $\mathcal{C}(X)$ to $\mathcal{C}(X)$, then P is called (weak) Feller. If the transition probability kernel P maps all bounded measurable functions to $\mathcal{C}(X)$, then P is called *strong* Feller.

Definition 10.0.5. A σ-finite measure π is *invariant* if

$$\pi(A) = \int_X \pi(dx)P(x, A), A \in \mathcal{B}(X)$$

.

Definition 10.0.6. A sequence of probabilities $\{\mu_k : k \in \mathbb{Z}_+\}$ is called *tight* if for each $\epsilon > 0$, there exists a compact subset $C \subset X$ such that

$$\liminf_{k\to\infty} \mu_k(C) \geq 1 - \epsilon$$

.

Definition 10.0.7. A chain $\mathbf{\Phi}$ is called *bounded in probability on avearge* if for any initial state $x \in X$, the sequence $\{\frac{1}{k}\sum_{i=1}^{k} P^i(x, \cdot) : k \in \mathbb{Z}_+\}$ is tight.

Theorem 10.0.1 (Theorem 12.0.1 (i) of Meyn and Tweedie (2005)). If $\mathbf{\Phi}$ is a weak Feller chain which is bounded in probability on average, then there exists at least one invariant probability measure.

Definition 10.0.8. A point $x \in X$ is called *reachable* if for every open set $O \in \mathcal{B}(X)$ containing x,

$$\sum_n P^n(y, O) > 0, \text{ for any } y \in X$$

.

Definition 10.0.9. The Markov transition function P is called *equicontinuous* if for each $f \in C_c(X)$, the sequence of functions $\{P^k f : k \in \mathbb{Z}_+\}$ is equicontinuous on compact sets. So the Markov chain which possesses an equicontinuous Markov transition function will be called an *e-chain*.

Proposition 10.0.1 (Proposition 6.4.2 of Meyn and Tweedie (2005))**.** *Suppse that the Markov chain Φ has the Feller property, and that there exits a unique probability measure π such that for every x, $P^n(x,\cdot) \xrightarrow{w} \pi$. Then Φ is an e-chain.*

Theorem 10.0.2 (Theorem 18.4.4 of Meyn and Tweedie (2005)). Suppose that Φ is an e-chain which is bounded in probability on average. Then a unique invariant probability π exists if and only if reachable state $x^* \in X$ exists.

Definition 10.0.10. A Markov chain $\{X_t\}$ is *ergodic* if there exists a probability measure ϕ such that

$$\sup_{A \in \mathcal{B}(X)} |P^n(x,A) - \pi(A)| \longrightarrow 0, asn \to 1$$

for any $x \in X$, where $P^n(x,A) = P(X_n \in A | x_0 = x)$.

Definition 10.0.11. If μ is a signed measure on $\mathcal{B}(X)$ then the *total variation norm* $||\mu||$ is defined as

$$||\mu|| := \sup_{f:|f|\leq 1} |\mu(f)| = \sup_{A \in \mathcal{B}(X)} \mu(A)$$

For an ergodic Markov chain, we have

$$\lim_{n\to\infty} ||P^n(x,\cdot) - \pi|| = 2 \lim_{n\to\infty} \sup_A |P^n(x,A) - \pi(A)| = 0$$

Definition 10.0.12. A Markov chain $\{X_t\}$ is *geometrically ergodic* if there exists $\rho \in (0,1)$ such that for any $x \in X$,

$$||P^n(x,\cdot) - \pi(\cdot)||_{TV} = o(\rho^n)$$

Theorem 10.0.3 (Theorem 6.2.3 in Liu (2012)). Suppose $\{X_t, t \geq 1\}$ is a Feller chain, and there exists a measure ϕ and a compact set A with $\phi(A) > 0$, such that
(i) $\{X_t\}$ is ϕ-irreducible,
(ii) there exists a non-negative continuous function g: $X \to \mathbb{R}^1$, such that

$$g(X) \geq 1, \ for\ all\ x \in A$$

and for some $\rho \in (0,1)$,

$$E[g(X_{t+1})|X_t = x] \leq (1-\rho)g(x), \ for\ x \in A^c$$

then $\{X_t\}$ is geometrically ergodic.

There are also some definitions and theorems about iterated random functions we have used to prove the propositions in Chapter 6 in Wu and Shao (2004). Following the notation in Wu and Shao (2004), let (X, τ) be the a complete separable metric space with Borel sets $\mathcal{B}(X)$, an iterated random function system on the state space X is defined as

$$X_n = F_{\theta_n}(X_{n-1}), \ n \in \mathbb{N}$$

where θ, θ_n, $n \in \mathbb{N}$, takes values in a second measurable space Θ, and are independent with identical marginal distribution H. Here, $F_\theta(\cdot) = F(\cdot, \theta)$ is the θ-section of a jointly measurable function $F : X \times \Theta \to X$ and X_0 is independent of $(\theta_n)_{n \geq 1}$.

Definition 10.0.13 (Wu and Shao (2004)). Geometric-Moment Contracting
Assume π is an invariant probability measure of the Markov chain $\{X_n\}$. Let $X_0' \sim \pi$ independent of $X_0 \sim \pi$ and $(\theta_k)_{k \geq 1}$. $X_n(X_0')$ can be viewed as a coupled version of $X_n(X_0)$, where $X_n(x) = F_{\theta_n} \circ F_{\theta_{n-1}} \circ \cdots \circ F_{\theta_1}(x)$. Then X_n is a *geometric-moment contracting* if there exists $\alpha > 0$, $C = C(\alpha) > 0$ and $r = r(\alpha) \in (0, 1)$, such that for all $n \in \mathbb{N}$,

$$E\{\tau^\alpha[X_n(X_0'), X_n(X_0)]\} \leq Cr^n$$

Condition 1 (Condition 1 in Wu and Shao (2004)). There exists $y_0 \in X$ and $\alpha > 0$ such that

$$I(\alpha, y_0) := E\{\tau^\alpha[y_0, F_\theta(y_0)]\} = \int_\theta \tau^\alpha[y_0, F_\theta(y_0)] H\{d\theta\} < \infty$$

Condition 2 (Condition 2 in Wu and Shao (2004)). There exists $x_0 \in X$, $\alpha > 0$, $r(\alpha) \in (0, 1)$ and $C(\alpha) > 0$ such that

$$E\{\tau^\alpha[X_n(x), X_n(x_0)]\} \leq C(\alpha) r^n(\alpha) \tau^\alpha(x, x_0)$$

Definition 10.0.14. A random variable is said to have an *algebraic tail* if there exists $A, B > 0$, such that $P(|Y| > y) < A/(B^y)$ for all $y > 0$.

Theorem 10.0.4 (Theorem 1 in Wu and Shao (2004) from Diaconis and Freedman

(1999)). Assume that condition 1 holds,

$$E(logK_\theta) = \int_\Theta logK_\theta H(d\theta) < 0, \ where \ K_\theta = \sup_{x' \neq x} \frac{\tau[F_\theta(x'), F_\theta(x)]}{\tau(x', x)}$$

and that K_θ has an algebraic tail. Then there exists a unique stationary distribution π for 10.0.13 and $Z_n(x) \to Z_\infty \sim \pi$ at a geometric rate. The limit Z_∞ dose not depend on x.

Theorem 10.0.5 (Theorem 2 in Wu and Shao (2004)). Suppose that Conditions 1 and 2 hold. Then there exists a random variable Z_∞ such that for all $x \in X$, $Z_n(x) \to Z_\infty$ almost surely. The limit Z_∞ is $\sigma(\theta_1, \theta_2, \cdots)$-measurable and does not depend on x. Moreover, for every $n \in \mathbb{N}$,

$$E\{\tau^\alpha[Z_n(x), Z_\infty]\} \leq C_r^n(\alpha)$$

where $C > 0$ depends solely on x, x_0, y_0 and α, and $r(\alpha) \in (0, 1)$. In addition, 10.0.13 holds.

Theorem 10.0.6 (Lemma A.1 in in Fokianos and Tjøstheim (2011)).

$$E(log(Y_t + 1)|v_t = v) - v \to 0 \quad as \quad v \to \infty$$

In this thesis, same as Liu (2012), the properties and theorems related to τ-weak dependence are also used to prove the propositions of ergodic property in Chapter 6. The definitions and theorems relate to the proof are given in following, which can

be referred to Doukhan and Wintenberger (2008) and Dedecker and Prieur (2004) for details.

Definition 10.0.15 (β-mixing coefficient in Dedecker and Prieur (2004)). Let (Ω, $\mathcal{A}$, $\mathbb{P}$) be a probability space, $\mathcal{M}$ a σ-algebra of $\mathcal{A}$ and X a real-valued random variable with distribution P_X, and $P_{X|\mathcal{M}}$ is a conditional distribution given $\mathcal{M}$, then the β-mixing coefficient between $\mathcal{M}$ and $\sigma(X)$ is defined as

$$\beta(\mathcal{M}, \sigma(X)) = \frac{1}{2}||V(P_{X|\mathcal{M}})||_1$$

where $||V(P_{X|\mathcal{M}})||_1 = \sup\left\{ |\int f(x)P_{X|\mathcal{M}}(dx) - \int f(x)P_X(dx)| : ||f_\infty|| < 1 \right\}$ is a $\mathcal{M}$-measurable random variable.

One of the most important properties of β is the coupling property proved by Berbee (1979): if Ω is rich enough, there is a random variable X^* independent of $\mathcal{M}$ and distributed as X such that $\mathbb{P}(X \neq X^*) = \beta(\mathcal{M}, \sigma(X))$.

Definition 10.0.16 (τ coefficient in Dedecker and Prieur (2004)). If the real-valued random variable X is integrable, the coefficient τ is defined by

$$\tau(\mathcal{M}, X) = ||W(P_{X|\mathcal{M}})||_1$$

where $||W(P_{X|\mathcal{M}})||_1 = \sup\left\{ |\int f(x)P_{X|\mathcal{M}}(dx) - \int f(x)P_X(dx)|, f \in< \Lambda_1(\mathbb{R}) \right\}$, and $\Lambda_1(\mathbb{R})$ is a 1-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$.

In Doukhan and Wintenberger (2008), $||\cdot||_m$ denotes $\mathbb{L}^m$-norm, i.e.,$||X||_m^m = \mathbb{E}||X||^m$ for $m \geq 1$, for every E-valued random variable X. For h: $E \to \mathbb{R}$, we define $||h||_\infty = \sup_{x \in E}|h(x)|$ and

$$\text{Lip}(h) = \sup_{x \neq y} \frac{h(x) - h(y)}{||x - y||}$$

.

The coupling argument also works on this τ coefficient:

$$\tau(\mathcal{M}, X) \leq ||X - Y||_1$$

for any Y with the same distribution as X and independent of $\mathcal{M}$, and this can be reached by some particular random variable X^* see Major (1978). In Doukhan and Wintenberger (2008), the dependence between the past of the sequence $(X_t)_{t \in \mathbb{Z}}$ and its future k-tuples may be assessed by the definition of τ coefficient: Consider the norm $||x - y|| = ||x_1 - y_1|| + \cdots + ||x_k - y_k||$ on E^k, set $\mathcal{M}_p = \sigma(X_t, t \leq p)$ and define

$$\tau_k(r) = \max_{1 \leq l \leq k} \frac{1}{l} \sup\{\tau(\mathcal{M}_p, (X_{j,1}, \cdot X_{j,l})) \quad \text{with} \quad p + r \leq j_1 < \cdots < j_l\}$$

$$\tau_\infty(r) = \sup_{k > 0} \tau_k(r)$$

For the sake of simplicity, $\tau_\infty(r)$ is denoted by $\tau(r)$. Finally, the time series $(X_t)_{t \in \mathbb{Z}}$ is $\tau - \textit{weakly dependent}$ when its coefficients $\tau(r)$ tend to 0 as r tends to infinity.

Combining the Theorem 3.1, Remark 3.1 and Corollary 3.1 in Doukhan and Wintenberger (2008), Liu (2012) gave a theorem on τ-weakly dependent stationary solution to Markov chain under the assumptions in Doukhan and Wintenberger (2008).

Theorem 10.0.7 (Theorem 6.4.1 in Liu (2012)). For the Markov chain $X_n(x) = F_{\theta_n} \circ F_{\theta_{n-1}} \circ \cdots \circ F_{\theta_1}(x)$, if for all $x, y \in X$,

$$E||f(x,\theta) - f(y,\theta)|| \leq a||(x-y)||$$

where $0 < a < 1$, and $\mu_1 = E||f(0,\theta)||_1 < \infty$, then there exists a τ-weakly dependent stationary solution $\{X_t\}$, such that $E||X_0|| < \infty$ and $\tau(r) \leq 2\mu_1(1-a)^{-1}a^r$ for $r \geq 1$. In addition, $\{X_t\}$ is the unique causal Bernoulli shift solution and is automatically an ergodic process.

References

Acharya, V., Das, S., Sundaram, R., 2006. Arbitrage-free pricing of credit derivatives with rating transitions. *Financial Analysts Journal* 58, 28-44.

Altman,E., 1998. The importance and subtlety of credit rating migration. *Journal of Banking and Finance* 22, 1231-1247.

Altman, E., Rijken, H., 2004. How rating agencies achieve rating stability. *Journal of Banking and Finance* 28, 2679-2714.

Andersen, P., Borgan, O., Grill, R., Keiding, N., 1995. Statistical Models Based on Counting Processes. Springer, New York.

A.P.Dempster; N.M.Laird; D.B.Rubin., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 39, No. 1. (1977), pp. 1-38.

Bangia, A., Diebold, F., Kronimus, A., Schagen, C., Schuermann, T., 2002. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance* 26, 445-474.

Barnard, G. A. 1959. Control charts and stochastic processes(with discussion). *J. Roy. Statist. Soc. Ser.* B 21, 239-271.

Berbee, H. 1979. Random walks with stationary increments and renewal theory. *Math. Cent. Tracts.*

Blume, M., Lim, F., MacKinlay, A., 1998. The declining credit quality of us corporate debt: myth or reality? *Journal of Finance* 53, 1389-1413.

Bruche, M., Gonzá lez-Aguado, C., 2010. Recovery rates, default probabilities, and the credit cycle. *Journal of Banking and Finance* 34, 754-764.

Cantor, R. Mann, C., 2003a. Measuring the Performance of Corporate Bond Ratings. *Moody's Special Comment.*

Cantor, R. Mann, C., 2003b. Are corporated bond ratings procyclical. *Moody's Special Comment.*

Chernoff, H., Zacks, S. 1964. Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Mathemat. Statist.* 35, 999-1018.

Christensen, J., Hansen, E., Lando, D., 2004. Confidence sets for continuous-time rating transition probabilities. *Journal of Banking and Finance* 28, 2575-2602.

Cox, D., 1981. Statistical analysis of time-series: some recent developments. *Scand. J. Statist.* 8 (2), 93âĂŞ115.

Das,S., Freed, L., Geng, G., Kapadia, N., 2006. Correlated default risk. *Journal of Fixed Income.* 16, 7-32.

Davis, R., Dunsmuir, W., Streett, S., DEC 2003. Observation-driven models for Poisson counts. *Biometrika* 90 (4), 777-790.

Davis, R., Rodriguez-Yam, G., Parameter- and Observation-Driven State Space Models. http://www.stat.columbia.edu/~rdavis/lectures/Cyprus1_04.

pdf

Davis, R., Liu, H., 2012. Theory and inference for a class of observation-driven models with application to time series of counts. Preprint, arXiv:1204.3915.

Dedecker, J. and Prieur, C. 2004. Coupling for τ-dependent sequences and applications. *Journal of Theoretical Probability* 17(4), 861-855.

Diaconis, P. and Ylvisaker, D. 1979. Conjugate Priors for Exponential Families. *Annals of Statistics* 7, 269-281.

Diaconis, P. and Freedman, D. 1999. Iterated random functions. *SIAM Review* 41, 45-76.

Doukhan, P. and Wintenberger, O. 2008. Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications* 118(11), 19972013.

Duffie, D., Eckner, A., Horel, G., Saita, L., 2009. Frailty correlated default. *Journal of Finance* 64, 2089-2123.

Egloff, D., Leippold, M., Vanini, P., 2007. A simple model of credit contagion. *Journal of Banking and Finance*. 31, 2475-2492.

Ferland, R., Latour, A., Oraichi, D., 2006. Integer-valued GARCH process. J. *Time Ser. Anal.* 27 (6), 923-942.

Fons, J., 2002. Understanding Moody's Corporate Bond Ratings and Rating Process. *Moody's Special Comment.*

Fokianos, K., Rahbek, A., Tjøstheim, D., 2009. Poisson autoregression. *J. Am.Statist. Assoc.* 104 (488), 1430-1439.

Fokianos, K., Tjøstheim, D., 2011. Log-linear poisson autoregression. *J. of Multivariate Analysis* 102 (3), 563-578.

Frey,R., McNeil,A., 2007. Dependence defaults in models of portfolio credit risk. *Journal of Risk* 6, 59-92.

Frydman, H., Schuermann, T., 2008. Credit rating dynamics and markov mixture models. *Journal of Banking and Finance* 32, 1062-1075.

Henderson, S. G., Matteson, D., Woodard, D., 2011. Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics* 5, 800-828.

Jarrow, R., Lando, D., Turnbull, S., 1997. A markov model for the term structure of credit risk spreads. *Review of Financial Studies* 10,481-523.

Jarrow, R., Lando, D., Turnbull, S., 1998. A markov chain model for valuing credit risk derivatives. *Journal of Derivatives* 6,97-108.

Kü chler, U., Sørensen, M., 1997. Exponential Families of Stochastic Processes. *Springer-Verlag, New York* .

Kedem, B., Fokianos, K., 2002. Regression models for time series analysis. *Wiley Series in Probability and Statistics.* Wiley-Interscience [John Wiley & Sons],Hoboken, NJ.

Lai, T., Liu, Tong., Xing, H., 2009. A bayesian Approach to Sequential Surveillance in Exponential Families. *Communications in Statistics–Theory and Methods* 38, 2958-2968.

Lai, T., Xing, H., 2010. Sequential Change-point Detection When the Pre-and Post-Change Parameters are Unknown. *Sequential Analysis* 29, 162-175.

Lai, T., Xing, H., 2008. A simple Bayesian Approach to Multiple Change-points. *Technical Report, Department of Statistics, Stanford University.*

Lando, D., 2000. Some elements of rating-based credit risk modeling. In: Jegadeesh. N., Tuckman, B. (Eds.), Advanced Fixed-Income Valuation Tools. John Wiley, Chichester,pp.193-215.

Lando, D., Skødeberg, T., 2002. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance* 26, 423-444.

Levine,Ross. 2001. International Financial Liberalization and Economic Growth. *Review of international economics* 9(4),688-702.

Liao,H., Chen,T., Lu,C., 2009. Bank credit risk and structural credit models: agency and information asymmetry perspectives. *Journal of Banking and Finance* 33, 1520-1530.

Liu, H., 2002. Some models for time series of counts. *PhD.thesis, Columbia University, Dept. of Statistics.*

Lorden, G., 1971. Procedures for Reacting to a Change in Distribution. *The Annals of Mathematical Statistics* 41,520-527.

Meyn, S. and Tweedie, R. (2005) Markov Chains and Stochastic Stability. *Cambridge University Press.*

Major, P. 1978 On the invariance principle for sums of identically distributed random variables. *Journal of Multivariate analysis* 8, 487-517.

Neumann, M. H., NOV 2011. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17 (4), 1268-1284.

Nickell,P., Perraudin, W., Varotto, S., 2000. Stability of rating transitions. *Journal of Banking and Finance* 24, 203-227.

Perron, Pierre. 1989. The Great Crash, The Oil Price Shock, and the Unit Root Hypothesis. *Econometric* 57: 1361-1401.

Page, E. S. 1954. Continuous inspection schemes. *Biometrika* 41, 100-114.

Pollak, M. 1985. Optimal Detection of a Change in Distribution. *Annual of Statistics* 18,1464-1469.

Pollak, M., Siegmund, D. 1991. Sequential detection of a change in a normal mean when the initial value is unknown. *Ann. Statist* 19, 394-416.

Roberts, S.W. 1966. A comparison of Some Control Chart Procedures. *Technometrics* 8, 411-430.

Rydberg, T.H., and Shephard, N. (2000), A Modeling Framework for the Prices and Times of Trades on the New York Stock Exchange, in W.J. Fitzgerald, R.L. Smith, A.T. Walden, P.V. Young (Eds), Nonlinear and Nonstationary

Signal Processing, *Isaac Newton Institute and Cambridge University Press*, Cambridge, 2000, pp.217-246.

Shiryaev, A.N. 1963. On Optimum Methods in Quickest Detection Problems. *Theory of Probability and Its Applications* 8, 22-46.

Shiryaev, A.N. 1978. Optimal Stopping Rules. *Springer-Verlag, New York.*

Siegmund, D., Venkatraman, E. S. 1995. Using the generalized likelihood ratio statistics for sequential detection of a change-point. *Ann. Statist* 23, 255-271.

Sacasa, Noel. 2008. Preventing future crises. *Finance and Development.*Dec, 2008.

Streett, S. 2000. Some observation driven models for time series of counts. *Ph.D. thesis, Colorado State University, Department of Statistics.*

Tsaig, Y., Levy, A., Wang, S., 2011. Analyzing the impact of credit migration in a portfolio setting. *Journal of Banking and Finance* 35, 3145-3157.

Tjøstheim, D., 2012. Some recent theory for autoregressive count time series. *TEST* 21 (3), 413-438.

Valentinyi-Endrész, Marianna, . 2004. Structural breaks and financial risk management. *MNB working paper.*

Weissbach, R., Walter, R., 2010. A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics* 155, 188-194.

Wolfgang Schmid., Dobromir Tzotchev., 2004. Statistical Surveillance of the Parameters of a One-Factor Cox-Ingersoll-Ross Model. *Sequential Analysis.* Vol. 23, No. 3, pp. 379-412, 2004.

Wu, W. and Shao, X. (2004) Limit theorems for iterated random functions. *Journal of Applied Probability*, 41, 425-436.

Xing, H., Sun, N., Chen, Y., 2012. Credit rating dynamics in the pressence of unknown structural breaks. *Journal of Banking and Finance* 36, 78-89.

Yao,Y. 1984. Estimation of a noisy discrete-time step functions Bayes and empirical Bayes approach. *Ann. Statist* 12, 1434-1447.

Zhu, F., 2012. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *J. of Mathematical Analysis and Applications* 389 (1), 58-71.