Estimation and Analysis of Multivariate Jump Diffusion Models with Jump Clustering and Contagion Effects

Mr. Yotsanan Simakorn

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Financial Engineering
Department of Banking and Finance
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2014

การประมาณค่าและการวิเคราะห์ของแบบจำลองการกระโดดดิฟฟิวชั่นหลายตัวแปรที่มีการกระโดดแบบเกาะกลุ่มและผลกระทบแพร่ระบาด

นายยศนันท์ สิมากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมการเงิน ภาควิชาการธนาคารและการเงิน
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557

Thesis Title	Estimation and Analysis of Multivariate Jump Diffusion Models with Jump Clustering and Contagion Effects
By	Mr. Yotsanan Simakorn
Field of Study	Financial Engineering
Thesis Advisor	Assistant ProfessorThaisiri Watewai, Ph.D.

Accepted by the Faculty of Commerce and Accountancy, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

_____________________________Dean of the Faculty of Commerce and Accountancy
(Associate ProfessorPasu Decharin, Ph.D.)

THESIS COMMITTEE

_____________________________Chairman
(Anant Chiarawongse, Ph.D.)

_____________________________Thesis Advisor
(Assistant ProfessorThaisiri Watewai, Ph.D.)

_____________________________Examiner
(Sira Suchintabandid, Ph.D.)

_____________________________External Examiner
(ProfessorMarc Stuart Paolella, Ph.D.)

ยศนันท์ สิมากร : การประมาณค่าและการวิเคราะห์ของแบบจำลองการกระโดดดิฟฟิวชั่นหลายตัวแปรที่มีการกระโดดแบบเกาะกลุ่มและผลกระทบแพร่ระบาด (Estimation and Analysis of Multivariate Jump Diffusion Models with Jump Clustering and Contagion Effects) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.ไทยศิริ เวทไว, 58 หน้า.

การเคลื่อนไหวที่รุนแรงในทิศทางลบของผลตอบแทนในช่วงวิกฤตเศรษฐกิจคือหนึ่งในหัวข้อที่สำคัญในโลกการเงิน เมื่อมีเหตุการณ์ที่คาดไม่ถึงเกิดขึ้นมักจะทำให้เกิดความเสียหายอย่างต่อเนื่องและแพร่ระบาดไปสู่ตลาดอื่นๆ งานวิจัยนี้พัฒนาแบบจำลองทางการเงินที่พยายามจะจับผลของการกระโดดแบบต่อเนื่องและการแพร่ระบาดนี้ แบบจำลองถูกพัฒนาให้รองรับกับข้อมูลสินทรัพย์จำนวนมาก หลักฐานเชิงประจักษ์จากการศึกษาบนข้อมูลรายสัปดาห์ของตลาดที่พัฒนาแล้วและตลาดที่กำลังพัฒนาตั้งแต่ปี 2542 ถึง 2557 พบว่าการกระโดดในตลาดที่พัฒนาแล้วมีค่าเฉลี่ยการกระโดดที่ติดลบอย่างชัดเจน ในขณะที่ในตลาดที่กำลังพัฒนานั้นมีการกระโดดทั้งบวกและลบ การกระโดดที่เกิดขึ้นในสองตลาดนั้นจะเพิ่มอัตราการกระโดดไปอีกเล็กน้อย การกระโดดในตลาดที่พัฒนาแล้วมีการแพร่ระบาดไปยังตลาดที่กำลังพัฒนาสูงกว่าที่ตลาดกำลังพัฒนาแพร่ระบาดมาสู่ตลาดที่พัฒนาแล้วถึงแม้ว่าอัตราการกระโดดแบบสุ่มนั้นจะไม่มีความสำคัญเชิงสถิติ การเคลื่อนไหวของผลตอบแทนและการกระโดดในตลาดที่พัฒนาแล้วมีความสอดคล้องกันในขณะที่ตลาดกำลังพัฒนามีการเคลื่อนไหวและการกระโดดที่ค่อนข้างเป็นอิสระต่อกัน การศึกษาบนข้อมูลรายวันของตลาดอเมริกาและลาตินอเมริกาตั้งแต่ปี 2542 ถึง 2558 บ่งบอกหลักฐานการมีอยู่ของการกระโดดอย่างต่อเนื่องและการแพร่ระบาดรวมไปถึงความอสมมาตร ตลาดอเมริกาตอบสนองต่อการเคลื่อนไหวที่คาดไม่ถึงในตลาดลาตินอเมริกามากกว่าตลาดลาตินอเมริกาตอบสนองต่อตลาดอเมริกา แบบจำลองนี้ยังให้การกระจายของอัตราการกระโดดเมื่อรู้ข้อมูลอดีตของผลตอบแทนซึ่งจะมีประโยชน์ต่อการใช้ในเรื่องการป้องกันความเสี่ยงและการจัดสรรสินทรัพย์

ภาควิชา	การธนาคารและการเงิน	ลายมือชื่อนิสิต	....
สาขาวิชา	วิศวกรรมการเงิน	ลายมือชื่อ อ.ที่ปรึกษาหลัก	....
ปีการศึกษา	2557		

5582260126 : MAJOR FINANCIAL ENGINEERING
KEYWORDS: MULTIVARIATE JUMP DIFFUSION MODELS / JUMP CLUSTERING / CONTAGION EFFECTS / EXPECTATION-MAXIMIZATION ALGORITHM

YOTSANAN SIMAKORN: Estimation and Analysis of Multivariate Jump Diffusion Models with Jump Clustering and Contagion Effects. ADVISOR: ASST. PROF.THAISIRI WATEWAI, Ph.D., 58 pp.

Extreme negative equity return across markets during financial crisis is one of the main concerned topics in the financial world. When such unexpected events occur, the adverse effect tends to trigger further losses and spreads over to other markets. This research develops a novel financial model that tries to capture these jump clustering and contagion effects. The model is developed to handle a large number of assets. Based on empirical evidence on developed and emerging markets from weekly data since 1999 to 2014, we find that jumps in developed markets have negative means whereas the means of jumps in emerging markets are mixed. Arrivals of jumps in both markets slightly increase their jump intensities. The jumps from the developed markets cause a higher contagion effect to the emerging markets than jumps from the emerging markets do to the developed markets although stochastic jump intensities are not statistically significant. The developed markets are more integrated among themselves whereas correlation among the emerging markets are low. Based on daily data of the United States and Latin America markets from 1999 to 2015, we find the evidences on jump clustering and jump contagion effects along with their jump asymmetries. The United States market reacts more to unexpected movements in the Latin America market than the Latin America does to the United States market. The model provides the distribution of the jump intensities given the past information on returns, which can be useful for applications in risk management and asset allocation.

Department:	Banking and Finance	Student's Signature	....
Field of Study:	Financial Engineering	Advisor's Signature	....
Academic Year:	2014		

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Asst. Prof. Thaisiri Watewai for the continuous support on my study and research. His patience, motivation, and immense knowledge have guided me throughout the research and writing of this thesis.

Besides my advisor, I would also like to thank other thesis committee: Dr. Anant Chiarawongse, Dr. Sira Suchintabandid, and Dr. Marc Stuart Paolella for their insightful comments and guidance.

I would like to thank Chulalongkorn Business School and its program, Master of Science in Financial Engineering, in providing the facilities and data resources in order to complete the research.

I thank the Capital Market Research Institute from the Stock Exchange of Thailand in providing the scholarship funding and comments on possible improvement of the research.

Lastly, I would like to thank my parents for their support and encouragement throughout all my studies at Chulalongkorn University.

CONTENTS

Page

1. Introduction

Poisson Jump Models are popular in option pricing and risk management due to the fact that they can capture the unexpected movement of the underlying asset's indexes being modeled. In other words, the models are capable of creating leptokurtic distribution which are normally found in financial asset returns distribution while ordinary diffusive models cannot. Until recently, most of the Poisson Jump Models are modeled based on the constant arrival rate of the jump or jump intensity. However, it has been found that asset's returns jump movement are having the properties called jump clustering and contagion effect. This has led to more studies toward conditional jump intensity models.

As the jumps are introduced into the models to solve the extreme cases of asset's indexes movement, it is clearly important that the jumps introduced should provide significant benefits to the model than without them. Furthermore, although in real life, jumps rarely happen, they still cause great impact as they would happen during crises. With the potential to create a huge amount of impact, it is essential to provide the jump characteristics that are similar to the real-life asset's jumps into the jump models.

Although some researchers have tried incorporating jump clustering and contagion effect into the models, there are some existing limitations in their models. This research will try to propose a multivariate jump diffusion models that could relax some of those limitations.

2. Background

2.1 Definitions

Different researches might define jump clustering and contagion effect differently. For this research, I define them like those as in the work of Aït-Sahalia, Cacho-Diaz, and Laeven (2014). In particular, jump clustering is the cluster of jumps in time series and the contagion effect is the contamination effect cross-sectionally towards other assets. For example, in terms of asset returns modeling, when a jump occurs, it changes the probability of the near next jump occurrence causing the clustering of jumps in the asset return. Further, these jumps can change the probability of another assets' jump occurrence. This effect is referred as the contagion effect.

2.2 Variations of Jump Models

The most well-known jump model would be the jump diffusion model introduced by Merton (1976). This type of model adds jump component to the diffusion model. However, there are variations in jump components and variations in the diffusion components. Some of the classification are being made in the survey by Sepp (2003) and those variations are stochastic volatility jump diffusion model, jump diffusion model with stochastic jump intensity, jump diffusion model with stochastic volatility and stochastic jump intensity, and jump diffusion model with deterministic volatility and jump intensity. Each variations also varies in the detailed components. Although, there are pure jump models, I do not consider that variation because I would like the jump component to act as the trigger to an unexpected event. This research would falls in the variation of jump diffusion model with stochastic jump intensity.

2.3 Variations of Jump Size

The most widely-used jump process is also from the model proposed by Merton and it is known as the compound Poisson process Merton (1976). This process has taken into account the amount of jumps from Poisson process and each of the jumps has a random jump size. Merton also has been using the randomization of jump size in log-normal form for asset's return. This is one of the most popular jump size used due to the fact that it is log-normally distributed like most of the return's asset distribution models so adding them up would create a nice form. Another popular type of jump size studied by Kou (2002) is double exponential jumps, which are more flexible than log-normal that it determines the heaviness of the left and right tails separately. Kou's model is demonstrated by Kou and Wang for an easy implementation on option pricing Kou and Wang (2004). In the survey done by Sepp (2003), there is also a weighted mixture of independent jumps where each independent jumps are given some amount of different weight for the jump as a whole similar to the compound Poisson process. This research will assume the log-normal return distribution.

2.4 Variations of Jump Intensity

Two of the most used variations of jump intensity in similar type of research are Cox process (also known as doubly stochastic intensity) introduced by Cox (1955) and Hawkes process introduced by Hawkes (1971). The Cox process is referred to the counting process where the jump intensity is some stochastic random variables. However, in some of the works like the works of Basu and Dassios (2002) and Dario and Simonis (2011), the jump intensity is not conditional on the prior jump intensity. Still, there are works which have taken account for prior intensity like the model studied by Fang (2000). The major difference between Cox process and Hawkes process mentioned by Aït-Sahalia et al. (2014) is that Cox process does not depend on the numbers of previous jumps while Hawkes process does.

This research will use the jump intensity that depends on the numbers of jumps in the prior period so it should fall in the category of a Hawkes process. In fact, we consider an extension of a Hawkes process.

2.5 Usefulness of EM algorithm

Couvreur (1997) describes the type of problem in maximum-likelihood estimation that can be benefitted with the help of the Expectation-Maximization algorithm or EM algorithm. EM algorithm is useful when we need to maximize the likelihood of the problems with variables that are unobservable or "incomplete data" problems. It is also useful in problems with many parameters to estimate. Couvreur (1997) states its advantages as simplicity and ease of implementation with the main drawback of possible slow convergence in some cases.

2.6 Concept of EM Algorithm

The main concept of the EM algorithm is to find the maximum likelihood estimators through the use of 2-steps iterative methods: Expectation (E-Step) and Maximization (M-Step). For easy implementation in these 2 steps, the EM algorithm has the concept called 'complete' and 'incomplete' data. Let Y be a random variable that is observable with probability density function $P(y|\Gamma)$ where Γ is the unknown parameters. Let Z be an unobservable variable or latent variable with probability density function $P(z|\Gamma)$. The 'complete' data refers to $\{Y, Z\}$ while the 'incomplete' data is $\{Y\}$. In the EM algorithm, instead of maximizing the likelihood function of the 'incomplete' data $L(\Gamma) = P(y|\Gamma)$, the algorithm first views the likelihood function of the 'complete' data $L_c(\Gamma) = P(y, z|\Gamma) = P(z|\Gamma)P(y|z, \Gamma)$.

In the E-Step, the expected complete-data log-likelihood $Q\left(\Gamma\middle|\Gamma^{(p)}\right) = \mathbb{E}\left[\ln L_c(\Gamma)\,\middle|\Gamma^{(p)}, y\right]$ is computed where p is the iteration number. In the M-Step, the parameter $\Gamma^{(p+1)}$ that maximizes $Q\left(\Gamma\middle|\Gamma^{(p)}\right)$ is determined; that is $\Gamma^{(p+1)} = \underset{\Gamma}{argmax}\, Q\left(\Gamma\middle|\Gamma^{(p)}\right)$. Each iteration is the alternating computation of the M-Step and E-Step until the convergence criteria are met. Under the regular condition, Γ converges to the maximum likelihood estimator.

3. Literature Review

3.1 Findings on Jump Effects

Many modelling researches have confirmed that there are effects on various assets' returns that provide characteristics like jump clustering effect and contagion effect. Chan and Maheu (2002) show in their work that having conditional jump intensity significantly outmatches constant jump intensity. Polson and Scott (2011)'s analysis show that most existing volatility models are unable to explain some features of contagion effects and there is a significant evidence on contagion effects during major EU crisis periods. Their model also suggests another effect called directional clustering in addition to the other two effects where they define them as the effect that the shock of aggregate volatility provides specific directional bias in the signs of country-level returns. Christoffersen, Kris, and Ornthanalai (2012) study using S&P500 index return and option data. They find evidence for time-varying jump intensities on both data, but option data's jumps are insignificant in low volatility regime. Additionally, they find that jump intensities seem to depend on market risk level. Choe and Lee (2013) test the models with various kinds of jump intensities and find consistent evidence on the existence of conditional asymmetry, in which they refer to as 'the difference between the dependence structures of up and down movements on past information'. Li and Zhang (2013) have additionally find that out-of-the-blue jumps correlate with stock price level historical average while the past jump intensity provides follow-on jumps (jump clustering) information. They also find that conditional expected jump size is negatively associated with stock price level historical average. Aït-Sahalia et al. (2014) show that the asset return model that considers jump clustering and contagion effect has provided evidence on self-excitation and reflecting across countries.

3.2 Findings on EM algorithm

Roche (2011) collects different variants of EM algorithm from various researchers. Two main classifications are either deterministic or stochastic. Most of the deterministic variants EM algorithm focus on speeding up algorithm by either simplifying computations or increasing the rate of convergence. Stochastic variants on the other hand focus on difficulties in implementation of either E-step or M-step in standard EM algorithm by replacing with stochastic simulation. Although, stochastic variants might take up more computation it does come with a nice trait that due to stochastic simulation, there is lesser tendency to be trapped in local maxima or saddle points.

3.3 Some Related Existing Models

Merton (1976)'s model provides one of the most adopted jump models. This research is also one of the extension to Merton's model so this section of literature review would focus on the works relating to the research's topic that are proposing the extension of Merton's model.

Chan and Maheu (2002) model their jumps similar to autoregressive model AR (1) and called it ARJI. They also have the extension of stochastic volatility using GARCH. Their work is an early development of jump model that provides insight to the studies of jump clustering. So their model has yet to take account for the contagion effect.

Aït-Sahalia et al. (2014) have created the multivariate jump diffusion process model with stochastic volatility and conditional jump intensity. Their stochastic volatility is based on Heston's Model and their jump intensity is based on Hawke's process, in which the jump size is double exponential. Their results show that there are contagion effects across regions around the world. They use GMM method for fitting the data. The benefits of their GMM are that the result's estimation of moments can be found in close-form, wrong assumption of parameters' distributions are not crucial with the known confidence level of distribution correctness, and the results can be interpreted easily with the separation of different orders of moments. However, their jump intensity rate dynamics cannot include some normal noises due to the fact that it might cause a negative value on their jump intensity rate. They also assume that the jump size of the jump intensity rate is both cross-sectionally and serially independent from each other. On multivariate case, they have to assume some jump sizes of their jump intensity rates to be 0 to still hold good tractability in their estimation while the other jump sizes of their jump intensity rates are assumed to be some constant parameters.

Polson and Scott (2011) study clustering, contagion, and directional effects on EU crisis periods. They define each correlation of stochastic volatility in terms of across time, cross-section, and upon volatility shocks to study the three effects. However, their model does not study these effects in terms of jumps so it is quite different from the model proposed by this research.

Duncan, Randal, and Thomson (2009) propose the EM algorithm to estimate the maximum likelihood of the multivariate jump diffusion model (a variation of Merton's model) with EM algorithm. Their work show the simplicity of implementation and also the better numerical properties compared to the conventional numerical optimization while also lessening the sensitivity of starting value choices.

Nevertheless, their work uses a constant jump intensity, which means the jump clustering and contagion effect are unaccounted for.

Some recent similar studies also extend the model for microstructure or high frequency data. Carlsson, Foo, Lee, and Shek (2007) model the bivariate Hawkes process for high frequency trade prediction. The model contains only the Hawkes process, which is used to create signal to trigger buy and sell. Bacry, Delattre, Hoffman, and Muzy (2013) study a new stochastic price model at tick-by-tick up to two assets at once. They used Hawkes process as a self and mutually exciting intensities of jumps. They also account for the Epps effect (the correlation of the increments in microscopic scales). Bacry and Jean-Francois (2013) then create the first model that accounts for market price microstructure that includes random time arrival of price moves, discrete price grid, high frequency mean reversion, and correlation functions at various time scales. Their work mainly focuses on the microstructure in which they account for buy and sell order execution effects. Fonseca and Zaartour (2014) later create the stock price model with Hawkes process as an extension to Bacry et al. (2013)'s work to study lead-lag correlation between two indexes in microstructure. Due to the fact, that the lead-lag correlation only happens when the prices move in the same direction, they assume that the two stocks would need to be positively correlated. Also, they only study the lead-lag correlation in pair of indexes using a bivariate model. This research would not deal with the microstructure data and will not discuss further about their work. One of the reasons this research avoids microstructure data is because in dealing with the contagion effects, it needs to take into account the asset returns from different regions in which each countries' market microstructure are not only different, but would also need to compensate on time lag, which can be difficult at the microscopic level.

3.4 Research Similarity

This research would be the extension to the work of Duncan et al. (2009). The main extension is to change the constant jump intensity into conditional jump intensity which would incorporate jump clustering and contagion effect. The conditional jump intensity would be then be similar to the jumps from the work of Aït-Sahalia et al. (2014), but this research proposes a model that is more applicable with multiple assets with the help of EM algorithm. The variants of EM algorithm are developed with the focus on easiness in implementation.

4. Objective

The objective of this research is to create a multivariate jump diffusion model incorporating jump clustering and contagion effects that can be fitted and is applicable in a wide range of applications.

4.1 Contributions

This research provides three main contributions. The first contribution is that the model relaxes the assumptions made from similar models. The main contributions in the model would come from an extension to the work of Aït-Sahalia et al. (2014) in three aspects. First, this research is modelling the jump intensity rates in logarithmic form meaning that it can account for noises and would still make the jump intensity rates non-negative. Second, this research is using the jump sizes of the jump intensity rates that are cross-sectionally correlated. Third, the jump sizes of the jump intensities are assumed to be multi-dimensional vectors of normal distributions instead of being the constant parameters.

The second contribution is that it provides an estimation method for multivariate jump diffusion model with conditional jump intensity that is applicable to multiple assets by using the EM algorithm. The EM algorithm is a model-specific implementation method that would require nontrivial work for the proposed model. From the process of EM algorithm, it also provides the filtered and smoothed probabilities of jumps and jump intensities that can be computed as a byproduct. Also, given the proposed model, all parameters are able to be obtained from the close-form solutions at each step of the EM algorithm, which would require small effort in computation.

The last contribution is that it allows the analysis of asymmetry in contagion effect across regions. Statistical tests can be done using the information matrix.

4.2 Possible Applications

This section lists some possible implementations from the research's model, however, these are not implemented in this research.

Portfolio Asset Allocation. With the model's additional consideration on risk involvement of jump clustering and contagion effects, the portfolio asset allocation's decision would be affected through the change of expected return and the contagion and jump clustering risk.

Strategic Trading. Jump clustering and contagion effect might contain lead or lag horizontally (across time) or cross-sectionally (across regions, industries, or considered groups), which may in turn provide a high probability of successful strategic trading.

Risk Management. Forecasting of risk might be available if the estimated parameters are good enough. Value-At-Risk will also change with jump clustering or contagion effects.

Pricing Derivatives. Some extensions to the model to include the risk neutral measure might be able to help price derivatives that account for jump clustering and contagion effects.

5. Methodology

5.1 Model

Incorporating the jumps into the model creates the discontinuity in the dynamics. By letting $X_{i,t}$ denote the return of asset i at time t, the discrete-time of return is then given as:

$$X_{i,t} = \mu_i + \sigma_i \Delta W_t + \sum_{k=1}^{K} \sum_{j=1}^{\Delta N_{k,t}} \theta_{i,k,j,t} \qquad i = 1, \dots, n \qquad (1)$$

The model is divided into two parts like Merton's model. The first part is the diffusion term consisting of the drift term (μ_i) and the volatility (σ_i) where W_t is a standard Brownian motion. This research will refer to this diffusion term as $z_{i,t} = \mu_i + \sigma_i(W_t - W_{t-1})$ allowing it to be normally distributed with mean μ_i and variance σ_i^2. Σ denotes the covariance matrix of $z_t = [\, z_{1,t}, \dots, z_{n,t}]$. When no jump occurs, equation(1) is left with only the diffusion term ($X_{i,t} = z_{i,t}$). This means that the return is normally distributed during normal period with no jump.

The second part is the jump term ($\sum_{k=1}^{K} \sum_{j=1}^{\Delta N_{k,t}} \theta_{i,k,j,t}$). K is the number of types of jumps. $\Delta N_{k,t}$ is the number of jumps of type k in the time interval $(t-1, t]$, which is assumed to be Poisson distributed with mean $\lambda_{k,t}$ where $\lambda_{k,t}$ is the jump intensity at time t of jump type k. $\theta_{i,k,j,t}$ is the random jump size of asset i from jump type k of each j jump at time t, which is assumed to be normally distributed with mean $\nu_{i,k}$ and variance $\delta_{i,k}^2$. Δ_k denotes the covariance matrix of $\theta_{k,j,t} = [\theta_{1,k,j,t}, \dots, \theta_{n,k,j,t}]$, which is independent of j as jumps from the same type are independent and identically distributed. When a jump of any type k occurs, the jump term will not be zero and this makes the return have fat-tailed distribution.

The jump intensity of jump process of type k at any time t is not constant and follows the following equation:

$$\lambda_{j,t+1} = e^{a_j + b_j \Delta V_t} \cdot \lambda_{j,t}^{c_j} \cdot \prod_{k=1}^{K} \prod_{l=1}^{\Delta N_{k,t}} e^{\eta_{j,k,l,t}} \qquad j = 1, \dots, K \qquad (2)$$

Equivalently, the log-intensity process follows the following equation:

$$\ln \lambda_{j,t+1} = a_j + b_j \Delta V_t + c_j \ln \lambda_{j,t} + \sum_{k=1}^{K} \sum_{l=1}^{\Delta N_{k,t}} \eta_{j,k,l,t} \qquad j = 1, \dots, K \qquad (3)$$

where V_t is a standard Brownian motion independent of W_t. Let $u_{j,t} = a_j + b_j(V_t - V_{t-1})$ be normally distributed with mean a_j and variance b_j^2. B denotes the covariance matrix of $u_t = [u_{1,t}, \dots, u_{K,t}]$. $u_{j,t}$ acts as the diffusion term of jump intensity. c_j is the weight coefficient given to the jump intensity of previous time step. If c_j is less than 1 when no jump of jump intensity occurs, the jump intensity tends to diminish to its mean value implying that jump intensity process has mean-reverting property. $\eta_{j,k,l,t}$ is the random jump size of intensity of type j from the l^{th} jump of type k at time t, which is assumed to be normally distributed with mean $\kappa_{j,k}$ and variance $\omega_{j,k}^2$. Ω_k denotes the covariance matrix of $\eta_{k,l,t} = [\eta_{1,k,l,t}, \dots, \eta_{K,k,l,t}]$, which is independent of l as jumps from the same type are independent and identically distributed. Assume that the log-intensities at time 1 or $\ln \lambda_1 \equiv [\ln \lambda_{1,1}, \dots, \ln \lambda_{K,1}]$ are jointly normally distributed with mean vector a_0 and covariance matrix B_0. $\sum_{k=1}^{K} \sum_{l=1}^{\Delta N_{k,t}} \eta_{j,k,l,t}$ is the jump term in jump intensity. $\eta_{j,k,l,t}$ represents the increase in the intensity of jump of type j when there is a jump from type k. For $j = k$, this $\eta_{j,j,l,t}$ represents the self-exciting effect leading to jump clustering. For $j \neq k$, $\eta_{j,k,l,t}$ represents the contagion effect that causes an alteration in the tendency to jump of a certain jump type j from jump type k.

5.2 Estimation

5.2.1 Expectation (E-Step)

The parameters needed to estimate include:

$$\Gamma = \{\mu, \Sigma, \nu, \Delta, \kappa, \Omega, a_0, B_0, a, B, c\}$$

The complete data C is given as:

$$C = \{\mathrm{X}, \mathrm{Z}, \Delta\mathrm{N}, \theta, \eta, \lambda\}$$

where each variable starts from time t = 1 until t = T.

X is the only observable data and is considered as incomplete data in this EM algorithm. The E-Step calculates: $Q\left(\Gamma \middle| \Gamma^{(p)}\right) = \mathbb{E}[\ln L_C | \Gamma^{(p)}, x_1, \dots, x_T]$ where x_t represents the observed vector of returns at time t. Let T denote the last time step, the complete-data log-likelihood is given by:

$$\ln L_C = -\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|B_0| - \frac{1}{2}(\ln\lambda_1 - a_0)' B_0^{-1}(\ln\lambda_1 - a_0)$$

$$-\frac{K(T-1)}{2}\ln 2\pi - \frac{T-1}{2}\ln|B|$$

$$-\frac{1}{2}\sum_{t=2}^{T}\left(\ln\lambda_t - a - c\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)' B^{-1}\left(\ln\lambda_t - a - c\ln\lambda_{j,t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)$$

$$+\sum_{t=1}^{T}\sum_{k=1}^{K}\left[-\lambda_{k,t} + \Delta N_{k,t}\ln\lambda_{k,t} - \ln\left(\Delta N_{k,t}!\right)\right]$$

$$+\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\left[-\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\left(\eta_{k,l,t-1} - \kappa_k\right)'\Omega^{-1}\left(\eta_{k,l,t-1} - \kappa_k\right)\right]$$

$$+\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\left[-\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|\Delta| - \frac{1}{2}\left(\theta_{k,l,t} - \nu_k\right)'\Delta^{-1}\left(\theta_{k,l,t} - \nu_k\right)\right]$$

$$-\frac{nT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(z_t - \mu)'\Sigma^{-1}(z_t - \mu)$$

The conditional expectation in E-Step is given by:

Let X denotes $x_1, \ldots, x_T$ and X_t denotes $x_1, \ldots, x_t$. The expected log-complete likelihood is

$$Q\left(\Gamma \middle| \Gamma^{(p)}\right) = -\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|B_0| - \frac{1}{2}\mathbb{E}[(\ln\lambda_1 - a_0)' B_0^{-1}(\ln\lambda_1 - a_0)|\mathrm{X}]$$
$$-\frac{K(T-1)}{2}\ln 2\pi - \frac{T-1}{2}\ln|B|$$
$$-\mathbb{E}[\frac{1}{2}\sum_{t=2}^{T}\left(\ln\lambda_t - a - c\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)' B^{-1}\left(\ln\lambda_t - a - c\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)|\mathrm{X}]$$
$$+\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{K} -\lambda_{k,t} + \Delta N_{k,t}\ln\lambda_{k,t} - \ln\left(\Delta N_{k,t}!\right) \middle| \mathrm{X}\right]$$
$$+\mathbb{E}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\left\{\begin{matrix} -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| \\ -\frac{1}{2}\left(\eta_{k,l,t-1} - \kappa_k\right)'\Omega^{-1}\left(\eta_{k,l,t-1} - \kappa_k\right)\end{matrix}\right\}\middle|\mathrm{X}\right]$$
$$+\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t}} -\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|\Delta| - \frac{1}{2}\left(\theta_{k,l,t} - \nu_k\right)'\Delta^{-1}\left(\theta_{k,l,t} - \nu_k\right)\middle|\mathrm{X}\right]$$
$$-\frac{nT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\mathbb{E}[\sum_{t=1}^{T}(z_t - \mu)'\Sigma^{-1}(z_t - \mu)\,|\mathrm{X}]$$

5.2.2 Optimization (M-Step)

The parameters' values that maximizes $Q\left(\Gamma \middle| \Gamma^{(p)}\right)$ in M-Step are given by:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[z_t|\mathrm{X}] \qquad \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[(z_t - \hat{\mu})(z_t - \hat{\mu})'|\mathrm{X}]$$

$$\hat{\nu}_k = \frac{\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t}\right)\middle|\mathrm{X}\right]}{\sum_{t=1}^{T}\mathbb{E}[\Delta N_{k,t}|\mathrm{X}]} \qquad \hat{\Delta}_k = \frac{\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t} - \hat{\nu}_k\right)\left(\theta_{k,1,t} - \hat{\nu}_k\right)'\middle|\mathrm{X}\right]}{\sum_{t=1}^{T}\mathbb{E}[\Delta N_{k,t}|\mathrm{X}]}$$

$$\hat{\kappa}_k = \frac{\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1}\right)\middle|\mathrm{X}\right]}{\sum_{t=2}^{T}\mathbb{E}[\Delta N_{k,t-1}|\mathrm{X}]} \qquad \hat{\Omega}_k = \frac{\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\mathrm{X}\right]}{\sum_{t=2}^{T}\mathbb{E}[\Delta N_{k,t-1}|\mathrm{X}]}$$

$$\hat{a}_0 = \mathbb{E}[\ln\lambda_1\,|\mathrm{X}] \qquad \hat{B}_0 = \mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|\mathrm{X}]$$

$$\hat{a} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}[u_t|\mathrm{X}] \qquad \hat{B} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}[(u_t - \hat{a})(u_t - \hat{a})'|\mathrm{X}]$$

$$\hat{c} = c^{(p)} + \mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})\,\middle|\mathrm{X}\right]^{-1}\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a})\,\middle|\mathrm{X}\right]$$

Note: $diag(a)$ means the diagonal matrix wose diagonal elements are the elements in a.

5.2.3 Auxiliary Probabilities

Let $Normal(x; a, B)$ denote the probability density function at x of the multivariate Normal random vector with mean vector a and covariance matrix B, and $Poisson(n; \lambda)$ the probability mass function at n of the Possion random variable with mean λ. The following probabilities are needed to compute the expectations of the complete-data log-likelihood in the expectation step.

Forward probability:

Let $\alpha(\lambda_t, \Delta N_t, x_t | \mathrm{X}_{t-1})$ denote the likelihood of $\lambda_t, \Delta N_t, x_t$ given $\mathrm{X}_{t-1} = \{x_1 \dots, x_{t-1}\}$, and $\alpha_t(a, b, c) \equiv \alpha_t(\lambda_t = a, \Delta N_t = b, x_t = c |\, \mathrm{X}_{t-1})$ where t is the time index. We have

$$\alpha_1(\lambda, \mathrm{n}, x_1) = P(\lambda_1 = \lambda, \Delta N_1 = n, x_1)$$

$$= Normal(\ln \lambda; a_0, B_0) \cdot \prod_{k=1}^{K} \{Poisson(\mathrm{n}; \lambda_{k,1})\}$$

$$\cdot Normal(x;\ \mu + \mathrm{n}\nu, \Sigma + \mathrm{n}\Delta)$$

$$\alpha_t(\lambda, \mathrm{n}, x_t) = P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}, x_t | \mathrm{X}_{t-1})$$

$$= \sum_{\lambda_{t-1}, \mathrm{m}} Normal(x_t; \mu + \mathrm{n}\nu, \Sigma + \mathrm{n}\Delta) \cdot \prod_{k=1}^{K} Poisson(\mathrm{n}_k; \lambda_k)$$

$$\cdot Normal(\ln \lambda\,; a + \ln \lambda_{t-1} + \mathrm{m}\kappa, B + \mathrm{m}\Omega) \cdot \frac{\alpha_{t-1}(\lambda_{t-1}, \mathrm{m})}{P(x_{t-1} | \mathrm{X}_{t-2})}, \qquad t > 1$$

2-Step probability:

Let $\zeta(\lambda_t, \Delta N_t, \lambda_{t+1} | \mathrm{X})$ denote the likelihood of $\lambda_t, \Delta N_t, \lambda_{t+1}$ given $\mathrm{X} = \{x_1 \dots, x_T\}$, and $\zeta_t(\mathrm{a}, \mathrm{b}, \mathrm{c}) \equiv \zeta_t(\lambda_t = \mathrm{a}, \Delta N_t = \mathrm{b}, \lambda_{t+1} = \mathrm{c} |\, \mathrm{X})$ where t is the time index. We have

$$\zeta_t(\lambda, \mathrm{n}, \mathrm{r}) = P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}, \lambda_{t+1} = \mathrm{r} | \mathrm{X})$$

$$= \left(\sum_{m}^{\infty} \gamma_{t+1}(r,m)\right) \cdot \left(\frac{Normal(\ln r\,; a + c\ln\lambda + n\kappa, B + n\Omega)\alpha_t(\lambda,n)}{\sum_{s,p} Normal(\ln r\,; a + c\ln s + p\kappa, B + p\Omega)\alpha_t(s,p)}\right)$$

Backward probability:

Let $\gamma(\lambda_t, \Delta N_t | X)$ denote the likelihood of $\lambda_t, \Delta N_t$ given $X = \{x_1 \dots, x_T\}$, and

$\gamma_t(a,b) \equiv \gamma_t(\lambda_t = a, \Delta N_t = b|\, X)$ where t is the time index. We have

$$\gamma_T(\lambda,n) = P(\lambda_T = \lambda, \Delta N_t = n|X) = \frac{\alpha_T(\lambda,n)}{\sum_{r,s}\alpha_T(r,s)}$$

$$\gamma_t(\lambda,n) = P(\lambda_t = \lambda, \Delta N_t = n|X) = \sum_{r}^{\infty} \zeta_t(\lambda,n,r), \qquad t < T$$

Probability of jumps given return:

Let $P(\Delta N_t = n|X)$ denote the probability of $\Delta N_t = n$ given $X = \{x_1 \dots, x_T\}$. We have

$$P(\Delta N_t = n|X) = \sum_{\lambda} P(\lambda_t = \lambda, \Delta N_t = n|X) = \sum_{\lambda} \gamma_t(\lambda,n)$$

6. Scope of the Project

6.1 Research

The research propose a model and its estimation in general form that can be used for studying various type of assets. Sample data-fitting implementations will be done with regional weekly stock index's return which can be used to analyze for existence of jump clustering and asymmetry in contagion effect across regions.

6.2 Data

The study will be divided into two parts. The first part uses weekly data on six countries data while the second part uses daily data on two continents data.

For the first part, due to the fact that different markets open and close at different time, this study will use the weekly US dollar MSCI gross return of different countries to limit the effect of time difference. The study data involve 3 developed countries and 3 Asian emerging countries since the start of January 1999 until the end of June 2014. The 6 countries are United States (US), Germany (DE), France (FR), Thailand (TH), Indonesia (ID), and Philippines (PH). The data are obtained from Bloomberg database.

For the second part, daily US dollar MSCI gross return of United States (US) and Latin America (LA) are used for the study. United States and Latin America are mostly in the same time zone so there is no need to adjust for the time lag. The data also start from January 1999 until the end of June 2015.

7. Empirical Result

7.1 Multivariate case on weekly data

The test is setup with two types of jumps: jumps that occur only in developed markets or developed jumps and jumps that occur only in emerging markets or emerging jumps. However, jumps in one region may increase the likelihood of jumps in the other region; that is, contagion effects across regions are allowed.

7.1.1 Data Summary

Table 1: Data summary statistics.

	Countries	Mean (%)	Standard Deviation (%)	Skewness	Excess Kurtosis	Maximum (%)	Minimum (%)
Developed	United States	0.0943	2.5848	-0.7829	6.4867	11.5827	-20.0473
	Germany	0.0849	3.6607	-0.7891	4.9126	15.2032	-26.0641
	France	0.0816	3.3785	-0.9195	6.3240	13.8786	-26.6867
Emerging	Thailand	0.2425	4.2405	-0.4726	4.1984	17.2644	-29.0259
	Indonesia	0.2985	4.9469	-0.2280	2.8554	21.5421	-26.8420
	Philippines	0.1433	3.5105	-0.3025	2.7574	15.2418	-20.8028

Table 1 provides the summary statistics of the weekly MSCI return data. The mean of the developed market return is lower than the mean of emerging markets. The risk or the volatility in the developed market is lower than in the emerging markets. Developed markets return are more negatively skewed and the tails are fatter. Emerging markets return outliers are more spread out.

7.1.2 Result and Discussion

The parameters obtained from running the EM algorithm for 100 iterations are in Table 2 and Table 3. Table 2 shows all the parameters related in the dynamics of the return while Table 3 shows all the parameters related in the dynamics of the jump intensity.

Table 2: Obtained parameters on return dynamics from 100 iterations. The parameters' name are highlighted in gray boxes while their values are in the white boxes on their right.

		US	DE	FR	TH	ID	PH
μ		0.0023	0.0063	0.0059	0.0011	0.0015	0.0066
Σ (standard deviation)		0.0192	0.0273	0.0257	0.0365	0.0410	0.0314
Σ (correlation)	US	1.0000	0.8080	0.8075	0.5253	0.4446	0.5079
	DE		1.0000	0.9326	0.5909	0.5178	0.5125
	FR			1.0000	0.5917	0.4825	0.5180
	TH				1.0000	0.5804	0.5586
	ID					1.0000	0.5710
	PH						1.0000
ν	Developed Jump	-0.0099	-0.0408	-0.0381			
	Emerging Jump				0.0122	0.0133	-0.0473
***Δ* (standard deviation)**		0.0324	0.0478	0.0430	0.0480	0.0626	0.0447
***Δ* (correlation)**	US	1.0000	0.6901	0.6941			
	DE		1.0000	0.9270			
	FR			1.0000			
	TH				1.0000	0.3818	0.0905
	ID					1.0000	0.2087
	PH						1.0000

Although data summary statistics from Table 1 shows that the mean of the return is lower in developed markets, the parameter μ (mean of the return diffusion term) from Table 2 shows that developed markets have higher diffusion mean. This might be due to the fact that emerging markets are more unpredictable and that the high mean in statistics are mostly the numbers from the jump term instead. The fact is also confirmed with the mean jump size (ν) in emerging markets that shows some positive numbers. The negative value of 'ν' on developed type of jump implies that the jumps captured from this run have negative expected jump sizes to the return whereas they are mixed in emerging type of jump. The volatilities from diffusion term and jump size are both higher in emerging markets and the jump sizes are less correlated among the countries than in the developed markets.

Table 3: Obtained parameters on jump intensity dynamics from 100 iterations.

		Developed Jump	Emerging Jump
$\boldsymbol{a_0}$		-0.1397	-0.1637
$\boldsymbol{B_0}$ **(standard deviation)**		0.0677	0.0780
$\boldsymbol{B_0}$ **(correlation)**	Developed Jump	1.0000	0.0740
	Emerging Jump		1.0000
$\boldsymbol{a}$		-0.1437	-0.1656
$\boldsymbol{B}$ **(standard deviation)**		0.0663	0.0760
$\boldsymbol{B}$ **(correlation)**	Developed Jump	1.0000	0.0748
	Emerging Jump		1.0000
$\boldsymbol{c}$		-0.0125	-0.0068
$\boldsymbol{\kappa}$	Developed Jump	0.0114	0.0058
	Emerging Jump	0.0038	0.0106
$\boldsymbol{\Omega}$ **Developed jump (standard deviation)**		0.0302	0.0337
$\boldsymbol{\Omega}$ **Developed Jump (correlation)**	Developed Jump	1.0000	0.0605
	Emerging Jump		1.0000
$\boldsymbol{\Omega}$ **Emerging jump (standard deviation)**		0.0304	0.0325
$\boldsymbol{\Omega}$ **Emerging Jump (correlation)**	Developed Jump	1.0000	0.0639
	Emerging Jump		1.0000

Table 3 shows the parameters from the log-intensity dynamic. The diffusion term of jump intensities are slightly different both in terms of mean and volatility between developed and emerging markets as can be seen in parameters a and B. Negative 'c' implies that jump intensity has a mean-reverting property. The contagion effect between the two types of jumps can be portrayed through the jump term of the jump intensity and can be analyzed in the parameter κ and Ω. Positive 'κ' shows that there are some self-exciting and contagion effects on the jump intensity. Ω shows a slightly more volatile jump size of jump intensity in emerging markets than in developed markets.

The EM algorithm in itself ensures that the likelihood increases in each iterations. Figure 1 plots the log-likelihood obtained from each iteration run. It can also be noticed that the log-likelihood converges really fast with 0.005% relative tolerance within the first 100 iterations.

Figure 1: Log-likelihood of 100 iterations from EM algorithm.

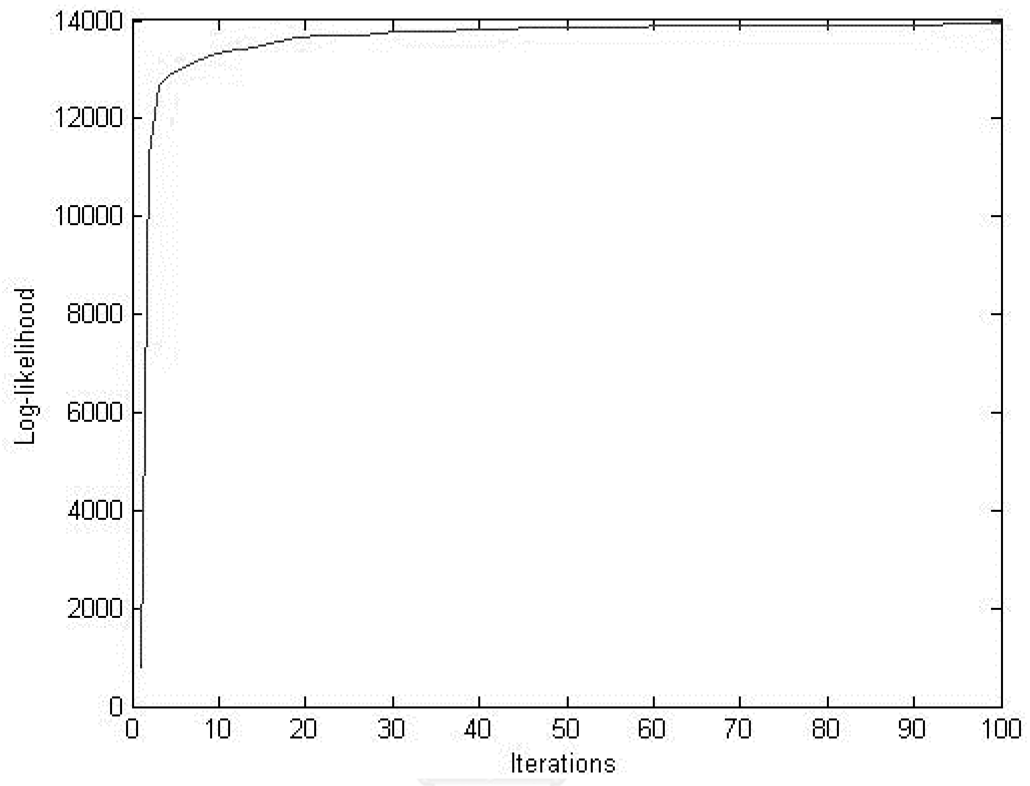


Figure 2: Return diffusion term for developed and emerging markets.

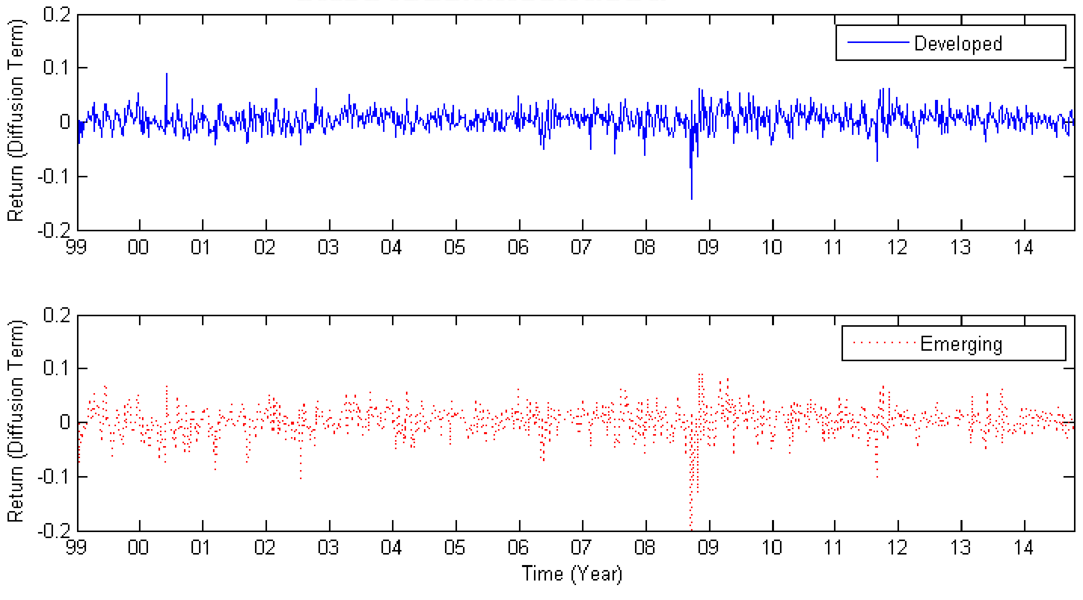


Figure 2 plots the smoothed expectation of the diffusion term of returns obtained from the model. The first plot (upper) shows the average of the diffusion term among the 3 developed countries and the second plot (lower) shows the average of the diffusion term among the 3 emerging countries. Most of the time, the graph shows the plot as a white noise, which is to be expected from the diffusion term. As stated before, the return diffusion term in emerging markets are more volatile than in developed countries as can be easily seen from the graph.

Figure 3 plots the smoothed expectation of the jump term of returns obtained from the average of 3 developed countries (upper plot) and the average of 3 emerging countries (lower plot). The graphs show that the first relatively large movement happens during 1999 to 2003. The second large movement starts from mid-2007 to almost 2010 and the third movement from 2010 to 2013 developed countries. The large jumps that are found during these three periods might come from the Dot-Com Bubble, the 2008 Great Financial Crisis, and European sovereign debt crisis. During 2003 to 2007, the graphs are almost flat. These graphs tell us that the jump term can capture some of the unexpected movements of the returns to a certain degree. Comparing the graph of return jump term between the developed jump type and emerging jump type from Figure 3 might not prove to be quite useful as it is the average plot. As stated earlier, the jumps of emerging markets are more mixed and unpredictable. Average plot of emerging jump type unfortunately cancels out making them look less volatile. Figure 9 in Appendix 2 shows each country' plots, which is easier to notice the unpredictive nature in emerging markets.

Figure 3: Return jump term for developed and emerging markets.

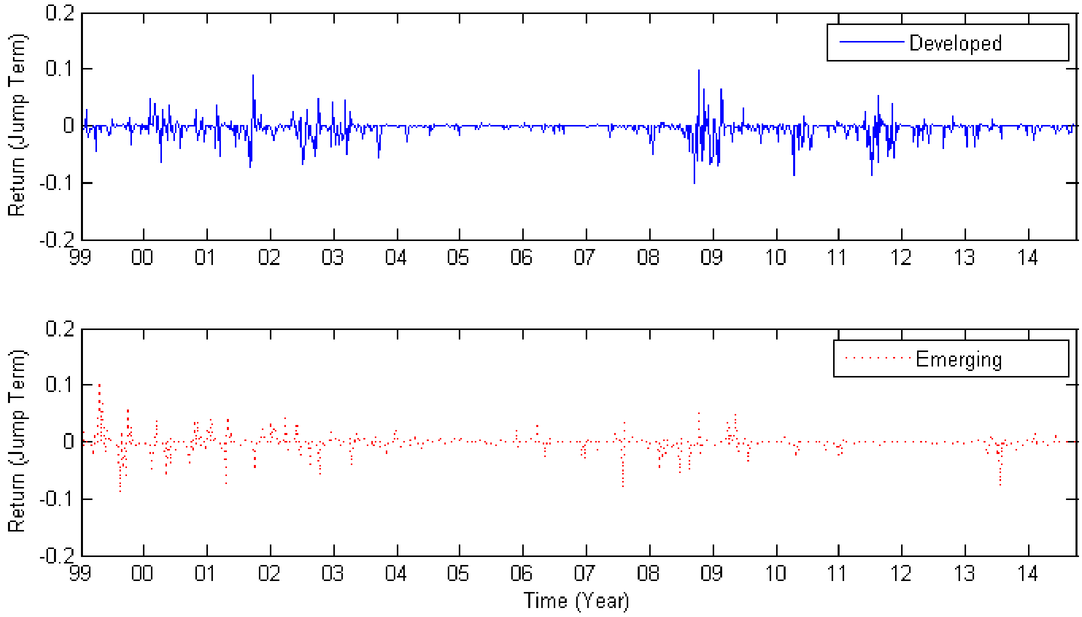


Figure 4: Expected jump intensity of each time-step t.

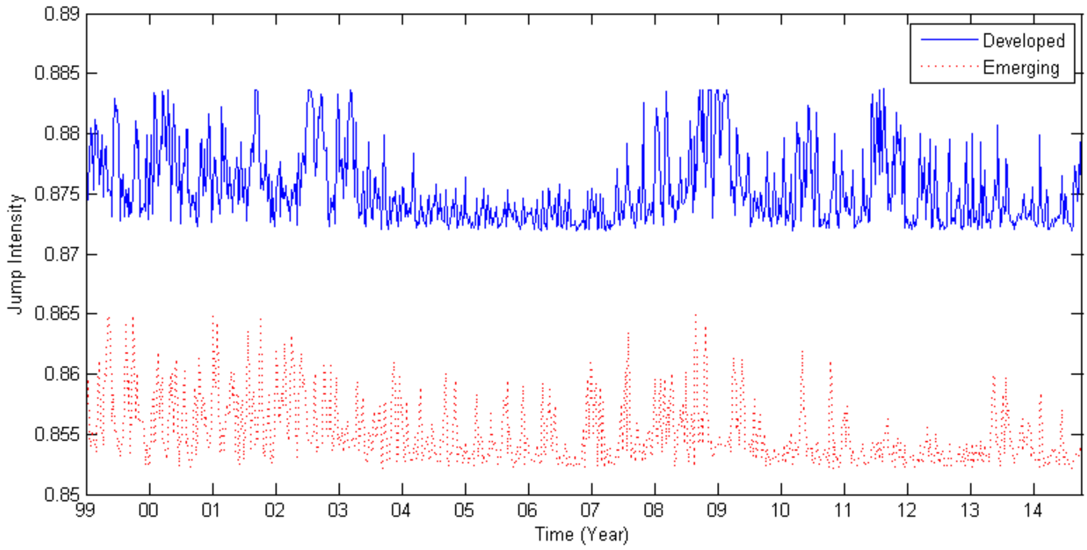


Figure 4 shows the movements of the expected jump intensity during 1999 to mid-2014. It shows higher rates during crises. After the rises of the rate, it stays high for a certain amount of time before settling down again. This is caused by the jump clustering effect and the mean-reverting property of the jump intensity in the model. It can also be seen that the rate increases significantly during mid-2007, which is around the start of the Global Financial Crisis. The graph also shows that emerging jump intensity is more random than in developed countries, which corresponds to the more unpredictive nature of the emerging markets.

Table 4: Average values of the parameters.

	Developed		Emerging
Return Diffusion Mean	0.48 %		0.31 %
Return Diffusion SD	2.41 %		3.63 %
Return Diffusion Correlation	0.85	0.52	0.57
Return Jumps Mean	-2.96 %		-0.73 %
Return Jumps SD	4.11 %		5.18 %
Return Jumps Correlation	0.77	0	0.23
Long-run Jump Intensity (Diffusion Term)	0.87		0.85
Jump Intensity Half-Life (week)	0.68		0.69

Table 4 shows the average values of the parameters on developed and emerging jump types. For correlation, the values are the average among their own type and the middle values show the average correlation between the two types. The result from the return diffusion term shows lower return with higher risk in emerging markets. Although this fact might be unsatisfactory to portfolio managers or investors who are interested in emerging markets, but with the much lower correlation of return diffusion, the emerging markets might have better diversification opportunities. Nevertheless, the higher volatility in emerging markets might outdo the diversification opportunities. In emerging markets, the return jump means are much lower in magnitude (some countries are actually positive). This tells us that there is less negative impact from emerging jumps compared to developed jumps. Still, the jumps in emerging markets are more unpredictable as can be seen in the return jump standard deviation. The return jump correlation also shows that developed markets are more systemic whereas emerging markets are more idiosyncratic. With the long-run jump intensity and their mean reversion rate that are only slightly different between the two market types, the jump terms might actually yield a better profit in emerging markets from similar number of jumps. Combining with the early strange fact of high risk, low return in emerging markets, this might prove to be that during normal time (diffusion term) or fundamental base of the emerging markets might be worse in investing, but during unexpected event (jump term) or speculative nature of the emerging market might create a higher fortune that comes with a greater risk involvement for speculative investing. This also aligns with the data summary statistics fact from Table 1 that emerging markets can provide higher return from higher risk.

Table 5: Jump clustering and contagion effects impact on other jumps.

Effects		To	
		Developed Jump	Emerging Jump
From	Developed Jump	1.25 %	0.68 %
	Emerging Jump	0.48 %	1.17 %

Table 5 shows the expected jump effects of a certain jump type to another jump type. When the effect is transferred from a jump type to the same jump type, it is a self-exciting effect that causes jump clustering. When the effect is transferred from a jump type to another jump type, it is the contagion effect. The number shows that developed jump self-excites the jump intensity by 1.25%. On average, the self-excitation seems to be small as the jump intensity is already high given the arrival of the first jump. The number thus shows the addition to the already high jump intensity of the last period. The contagion effect from developed jump type to emerging jump type is more than emerging jump type to developed jump type. This makes sense since developed countries should affect more to emerging countries than the other way around. Emerging jump also self-excites itself less than developed jump self-excites itself as emerging markets are less integrated among themselves.

Table 6: Parameters statistical significance on two-sided test.

Parameters	Value (Standard Error)	P-Value	Parameters	Value (Standard Error)	P-Value
μ_{US}	0.0023 (0.0008)	0.0041*	$\Delta_{US,DE}$	0.0011 (0.0001)	0.0000*
μ_{DE}	0.0063 (0.0011)	0.0000*	$\Delta_{US,FR}$	0.001 (0.0001)	0.0000*
μ_{FR}	0.0059 (0.0011)	0.0000*	$\Delta_{DE,DE}$	0.0023 (0.0002)	0.0000*
μ_{TH}	0.0011 (0.0016)	0.4893	$\Delta_{DE,FR}$	0.0019 (0.0002)	0.0000*
μ_{ID}	0.0015 (0.0018)	0.3822	$\Delta_{FR,FR}$	0.0018 (0.0002)	0.0000*
μ_{PH}	0.0066 (0.0014)	0.0000*	$\Delta_{TH,TH}$	0.0023 (0.0002)	0.0000*
$\Sigma_{US,US}$	0.0004 (0.0000)	0.0000*	$\Delta_{TH,ID}$	0.0011 (0.0002)	0.0000*
$\Sigma_{US,DE}$	0.0004 (0.0000)	0.0000*	$\Delta_{TH,PH}$	0.0002 (0.0002)	0.2551
$\Sigma_{US,FR}$	0.0004 (0.0000)	0.0000*	$\Delta_{ID,ID}$	0.0039 (0.0004)	0.0000*
$\Sigma_{US,TH}$	0.0004 (0.0000)	0.0000*	$\Delta_{ID,PH}$	0.0006 (0.0002)	0.0069*
$\Sigma_{US,ID}$	0.0003 (0.0000)	0.0000*	$\Delta_{PH,PH}$	0.0020 (0.0002)	0.0000*
$\Sigma_{US,PH}$	0.0003 (0.0000)	0.0000*	$\kappa_{DM,DM}$	0.0114 (76.928)	0.9999
$\Sigma_{DE,DE}$	0.0007 (0.0000)	0.0000*	$\kappa_{DM,EM}$	0.0058 (87.0315)	0.9999
$\Sigma_{DE,FR}$	0.0007 (0.0000)	0.0000*	$\kappa_{EM,DM}$	0.0038 (177.8009)	1.0000
$\Sigma_{DE,TH}$	0.0006 (0.0000)	0.0000*	$\kappa_{EM,EM}$	0.0106 (177.5716)	1.0000

Table 6 (continued) : Parameters statistical significance on two-sided test.

Parameters	Value (Standard Error)	P-Value	Parameters	Value (Standard Error)	P-Value
$\Sigma_{DE,ID}$	0.0006 (0.0000)	0.0000*	$\Omega_{DM-DM,DM}$	0.0009 (97.1534)	1.0000
$\Sigma_{DE,PH}$	0.0004 (0.0000)	0.0000*	$\Omega_{DM-DM,EM}$	0.0001 (139.183)	1.0000
$\Sigma_{FR,FR}$	0.0007 (0.0000)	0.0000*	$\Omega_{DM-EM,EM}$	0.0011 (245.3046)	1.0000
$\Sigma_{FR,TH}$	0.0006 (0.0000)	0.0000*	$\Omega_{EM-DM,DM}$	0.0009 (95.4296)	1.0000
$\Sigma_{FR,ID}$	0.0005 (0.0000)	0.0000*	$\Omega_{EM-DM,EM}$	0.0001 (143.1178)	1.0000
$\Sigma_{FR,PH}$	0.0004 (0.0000)	0.0000*	$\Omega_{EM-EM,EM}$	0.0011 (257.1241)	1.0000
$\Sigma_{TH,TH}$	0.0013 (0.0001)	0.0000*	$a_{0_{DM}}$	-0.1397 (894.808)	0.9999
$\Sigma_{TH,ID}$	0.0009 (0.0001)	0.0000*	$a_{0_{EM}}$	-0.1637 (735.9578)	0.9998
$\Sigma_{TH,PH}$	0.0006 (0.0000)	0.0000*	$B_{0_{DM}}$	0.0046 (850.5554)	1.0000
$\Sigma_{ID,ID}$	0.0017 (0.0001)	0.0000*	$B_{0_{DM,EM}}$	0.0004 (953.2893)	1.0000
$\Sigma_{ID,PH}$	0.0007 (0.0000)	0.0000*	$B_{0_{EM}}$	0.0061 (1518.0185)	1.0000
$\Sigma_{PH,PH}$	0.0010 (0.0000)	0.0000*	a_{DM}	-0.1437 (22.5321)	0.9949
ν_{US}	-0.0099 (0.0023)	0.0000*	$a_{DM,EM}$	-0.1656 (217.2242)	0.9994
ν_{DE}	-0.0408 (0.0037)	0.0000*	B_{DM}	0.0044 (14.2018)	0.9998
ν_{FR}	-0.0381 (0.0034)	0.0000*	$B_{DM,EM}$	0.0004 (672.8514)	1.0000
ν_{TH}	0.0122 (0.0041)	0.0029*	B_{EM}	0.0058 (257.9936)	1.0000
ν_{ID}	0.0133 (0.0049)	0.0069*	c_{DM}	-0.0125 (39.6979)	0.9997
ν_{PH}	-0.0473 (0.0042)	0.0000*	c_{EM}	-0.0068 (32.8074)	0.9998
$\Delta_{US,US}$	0.001 (0.0001)	0.0000*			

**Significance to 1% confidence level.*

The statistical test result in Table 6 shows that the parameters involving in the dynamics of the return are significant at 1% significant level while the parameters involving the dynamics of the jump intensity are rather insignificant. This might tell that the jumps of return are presented, however, the stochastic jump intensity might not be required. Still, there are some insignificant value of the parameters in the dynamics of the return like the μ_{TH}, μ_{ID}, and $\Delta_{TH,PH}$. This might means that this sample study might not correctly capture the jumps of certain emerging countries causing all jump intensity's parameters to be insignificant.

Table 7: Jump effects asymmetries statistical signficance on one-sided test.

Interested Asymmetries	Value (Standard Error)	P-Value
$\kappa_{DM,EM} > \kappa_{EM,DM}$	0.0019 (151.6336)	0.5000
$\kappa_{DM,DM} > \kappa_{DM,EM}$	0.0057 (93.6454)	0.5000
$\kappa_{DM,DM} > \kappa_{EM,DM}$	0.0076 (208.5249)	0.5000
$\kappa_{EM,EM} > \kappa_{DM,EM}$	0.0048 (155.4664)	0.5000
$\kappa_{EM,EM} > \kappa_{EM,DM}$	0.0068 (17.3127)	0.4998

The statistical test result in Table 7 shows that there are no statistical significance evidences on contagion asymmetries or that jump clustering are stronger than contagion effects. Due to the fact that parameters involving the study of the jump effects are not significant statistically, jump effect asymmetries are not able to be concluded.

7.2 Bivariate case on daily data

7.2.1 Data Summary

Table 8: Data summary statistics.

Data	***Mean (%)***	***Standard Deviation (%)***	***Skewness***	***Excess Kurtosis***	***Maximum (%)***	***Minimum (%)***
United States	0.0194	1.2423	-0.1908	8.2015	11.0426	-9.5039
Latin America	0.0414	1.7359	-0.3426	9.2237	15.3640	-15.0601

Table 1 provides the summary statistics of the daily MSCI return data. The mean of the US market return is lower than the mean of LA market. The volatility in the US market is lower than in the LA market. The LA market return is more negatively skewed, the tails are fatter, and the return outliers are more spread out.

7.2.2 Result and Discussion

The parameters obtained from running the EM algorithm for 30 iterations are in Table 9 and Table 10. Table 9 shows all the parameters related in the dynamics of the return while Table 10 shows all the parameters related in the dynamics of the jump intensity.

Table 9: Obtained parameters on return dynamics from 30 iterations.

	US	LA
$\boldsymbol{\mu \times 10^{-4}}$	0.4744	5.9485
Σ (standard deviation)	0.0111	0.0158
Σ (correlation)	0.7150	
$\boldsymbol{\nu}$	0.0093	-0.0136
***Δ* (standard deviation)**	0.0285	0.0382

Table 10: Obtained parameters on jump intensity dynamics from 30 iterations.

		US Jump	LA Jump
a_0		-3.5858	-3.6324
B_0 (standard deviation)		0.0551	0.0769
B_0 (correlation)		0.4474	
a		-3.5856	-3.6356
B (standard deviation)		0.0550	0.0768
B (correlation)		0.4475	
$c \times 10^4$		0.5278	-9.1276
κ	US Jump	0.0045	0.0007
	LA Jump	0.0103	0.0111
Ω US jump (standard deviation)		0.0290	0.0407
Ω US Jump (correlation)		0.6202	
Ω LA jump (standard deviation)		0.0286	0.0399
Ω LA Jump (correlation)		0.6153	
Long-run Jump Intensity		0.0265	0.0277
Jump Intensity Half-Life (day)		0.6925	0.6932

Table 9 shows that the mean of the diffusion is higher for Latin America with higher volatility. The correlation between United States and Latin America is high while the jump size in return is highly volatile. The US return has positive jump size mean while the LA return has negative jump size mean with higher volatility.

Table 10 shows quite high correlation between the US jump intensity and LA jump intensity on both B and Ω. The jump size of jump intensity κ shows that LA jump type provides high likelihood for the next jump to the US and LA markets than the US jump type. The long-run jump intensity and half-life are slightly different between the two jump types.

Figure 5 plots the log-likelihood obtained from each iteration run on EM algorithm. It can also be noticed that the log-likelihood converges really fast within the first 30 iterations.

Figure 5: Log-likelihood of 30 iterations from EM algorithm.

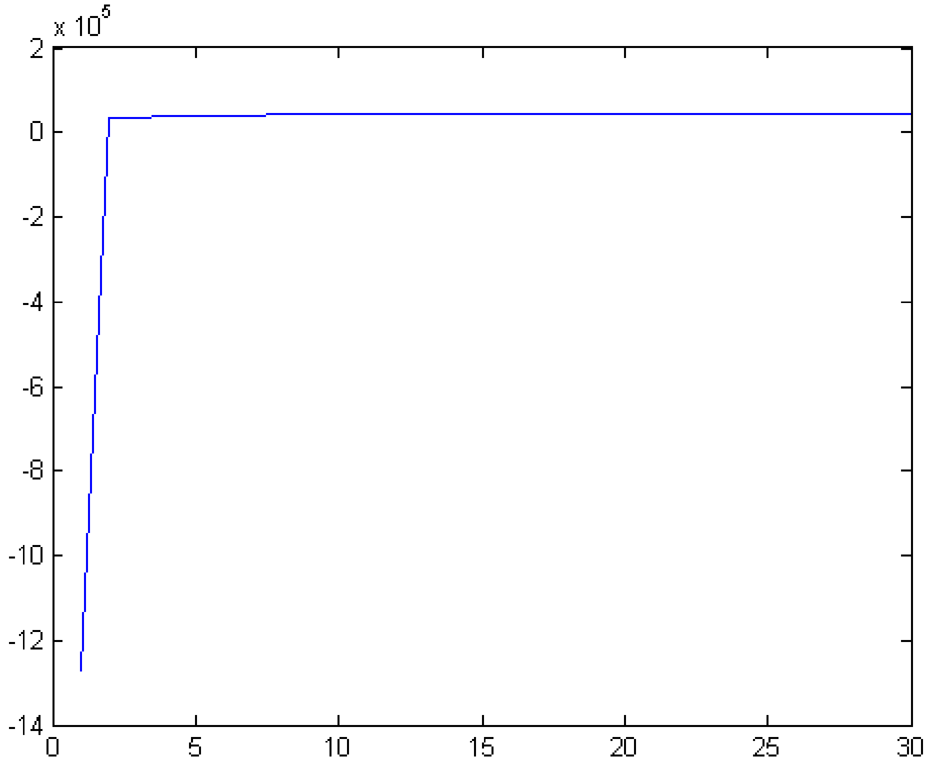


Figure 6: Return diffusion term for US and LA.

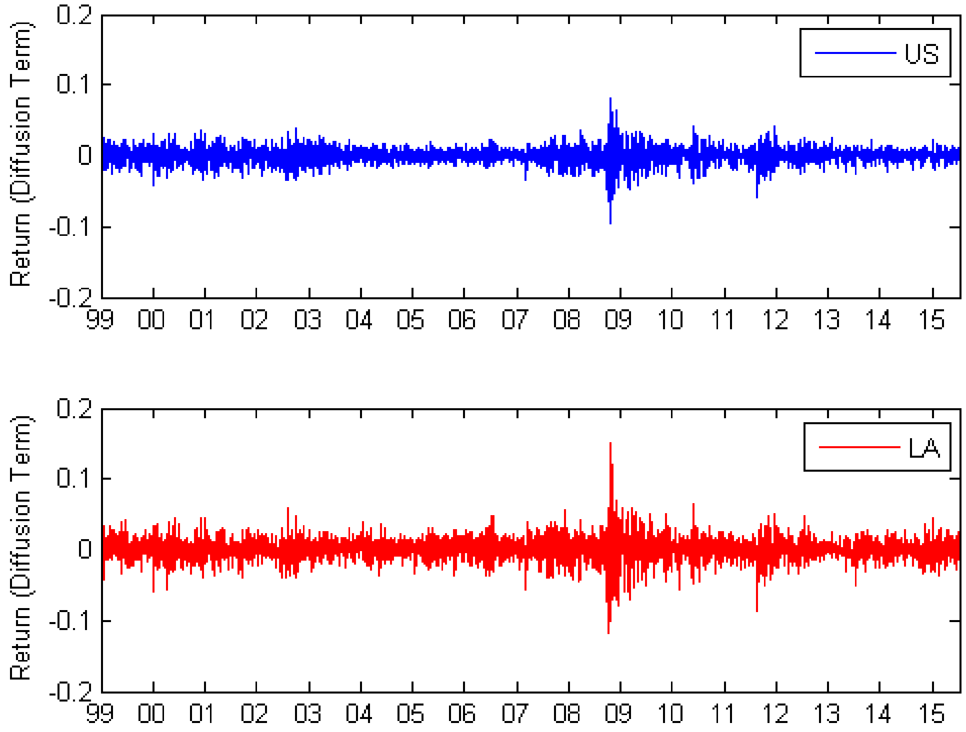


Figure 6 plots the smoothed expectation of the diffusion term of returns obtained from the model. The first plot (upper) shows the diffusion of the US market and the second plot (lower) shows the diffusion term of the LA market. During 2008, it seems that the jump term might not be able to capture all the high magnitude return from the high varying jump size (ν) requirement during the period reflecting in high diffusion.

Figure 7: Return jump term for US and LA.

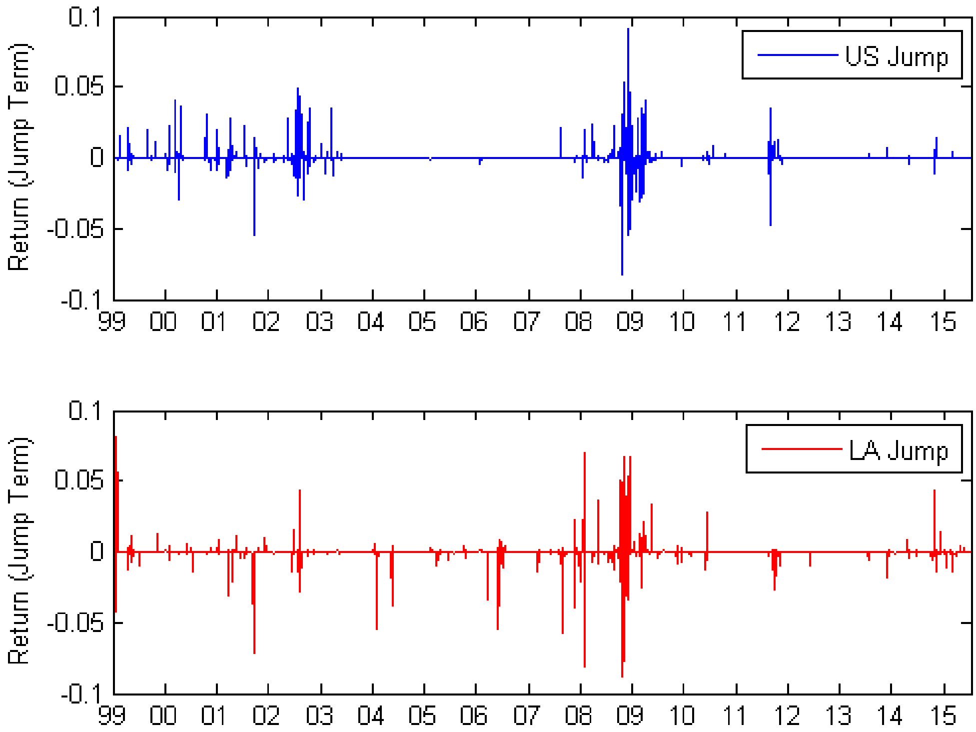


Figure 7 plots the smoothed expectation of the jump term of returns obtained from the US market (upper plot) and the LA market (lower plot). The graph shows similar result with the multivariate in Section 7.1 that it seems to be able to capture the jumps during the three financial crises: Dot-Com Bubble, the 2008 Great Financial Crisis, and European sovereign debt crisis.

Figure 8: Expected jump intensity of each time-step t.

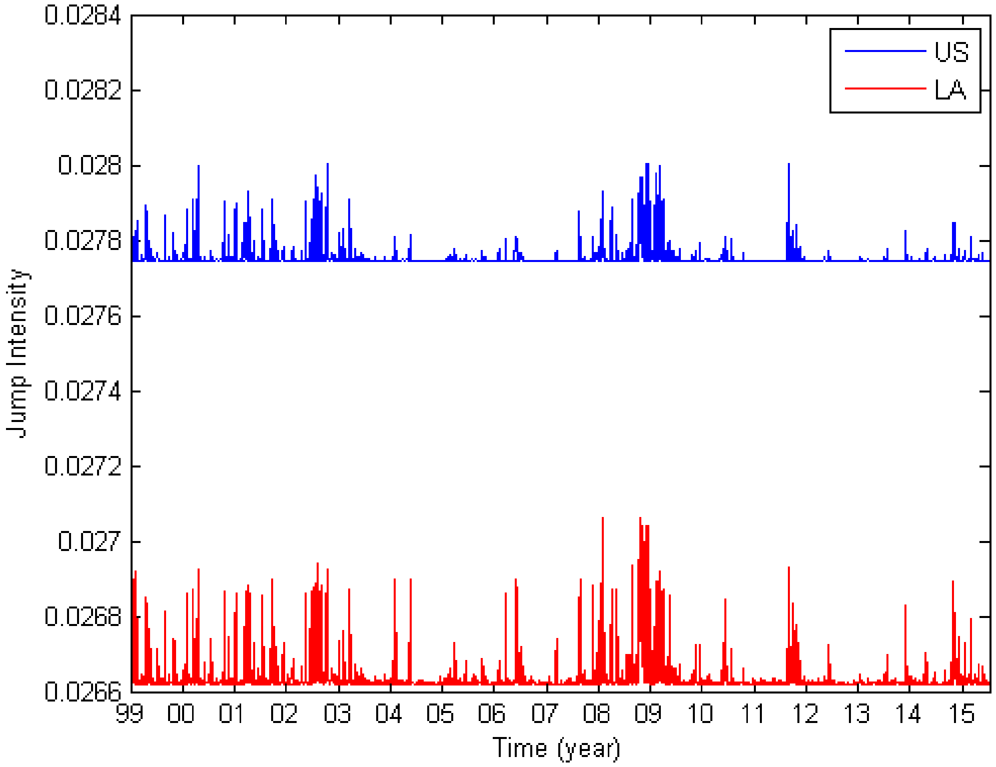


Figure 8 shows the movements of the expected jump intensity during 1999 to mid-2015. It shows higher rates during crises period. The LA intensity shows more jumps of itself during other periods than the US jumps.

Table 11: Jump clustering and contagion effects impact on other jumps.

Effects		To	
		US Jump	LA Jump
From	US Jump	0.5511%	0.2100%
	LA Jump	1.1368%	1.2566%

Table 11 shows the expected jump effects of a certain jump type to another jump type. The effect from the LA jump contributes largely to both US and LA jumps. The self-exciting effect in the US shows a lot higher than the contagion while the contagion effect in from the LA market is high and is almost equal to its self-exciting effect. This implies that the US market seems to react to most unexpected movements occurring in the LA market while the LA market does not react much to the US market.

Table 12: Parameters statistical significance on two-sided test.

Parameters	Value (Standard Error) $\times 10^{-4}$	P-Value	Parameters	Value (Standard Error) $\times 10^{-4}$	P-Value
μ_{US}	0.4744 (0.0000)	0	$a_{0_{US}}$	-35857.9457 (1.9602)	0
μ_{LA}	5.9485 (0.0003)	0	$a_{0_{LA}}$	-36323.9296 (1.9857)	0
$\Sigma_{US,US}$	1.2283 (0.0001)	0	$B_{0_{US}}$	30.3265 (0.0017)	0
$\Sigma_{US,LA}$	1.2518 (0.0001)	0	$B_{0_{US,LA}}$	18.9442 (0.001)	0
$\Sigma_{LA,LA}$	2.4959 (0.0001)	0	$B_{0_{LA}}$	59.1269 (0.0032)	0
ν_{US}	92.55 (0.0051)	0	a_{US}	-35856.0675 (2.1602)	0
ν_{LA}	-135.7209 (0.0074)	0	$a_{US,LA}$	-36356.3537 (2.0015)	0
Δ_{US}	8.122 (0.0004)	0	B_{US}	30.2329 (0.0016)	0
Δ_{LA}	14.5716 (0.0008)	0	$B_{US,LA}$	18.889 (0.001)	0
			B_{LA}	58.9283 (0.0032)	0
			c_{US}	0.5278 (0)	0
			c_{LA}	-9.1276 (0.0005)	0
			$\kappa_{US,US}$	44.9446 (0.0025)	0
			$\kappa_{US,LA}$	6.9607 (0.0004)	0
			$\kappa_{LA,US}$	103.1982 (0.0056)	0
			$\kappa_{LA,LA}$	111.0987 (0.0061)	0
			$\Omega_{US-US,US}$	8.4054 (0.0005)	0
			$\Omega_{US-US,LA}$	7.3183 (0.0004)	0
			$\Omega_{US-LA,LA}$	16.5644 (0.0009)	0
			$\Omega_{LA-US,US}$	8.1674 (0.0004)	0
			$\Omega_{LA-US,LA}$	7.017 (0.0004)	0
			$\Omega_{LA-LA,LA}$	15.9239 (0.0009)	0

The statistical test result in Table 12 shows that the parameters involving in the dynamics of the return are significant at 1% significant. Significance in $\kappa_{US,US}$ and $\kappa_{LA,LA}$ shows the evidence of the self-exciting effect while significance in $\kappa_{US,LA}$ and $\kappa_{LA,US}$ shows the evidence of the contagion effects. The test uses the covariance matrix of the parameter estimates from the approximation of the information matrix. The calculation detail can be found in Appendix 3.

Table 13: Jump effects asymmetries statistical signficance on one-sided test.

Interested Asymmetries	Value (Standard Error) $\times 10^{-4}$	P-Value
$\boldsymbol{\kappa_{US,LA} > \kappa_{LA,US}}$	-96.2375 (0)	0
$\boldsymbol{\kappa_{US,US} > \kappa_{US,LA}}$	37.984 (0)	0
$\boldsymbol{\kappa_{US,US} > \kappa_{LA,US}}$	-58.2536 (0)	0
$\boldsymbol{\kappa_{LA,LA} > \kappa_{US,LA}}$	104.138 (0)	0
$\boldsymbol{\kappa_{LA,LA} > \kappa_{LA,US}}$	7.9005 (0)	0

The statistical test result in Table 13 shows that all asymmetries are significant. Negative values means that the interested asymmetries are on the opposite direction. The first asymmetry shows that the contagion from LA to US is greater than US to LA. The second and third asymmetries show that the self-exciting effect in US is greater than the contagion from US to LA, but lower than the contagion from LA to US. The fourth and fifth asymmetries show that the self-exciting effect in LA is greater than the contagion effects from both ways.

8. Conclusion

This research develops a new financial model and its estimation method that can handle a large number of assets and allows for self-exciting and contagion effects with stochastic jump intensity. The run on the model has good convergence rate on both multivariate and bivariate cases.

In multivariate case, the model is able to capture large movements of return through the jumps from conditional jump intensity and tends to show that jump clustering and contagion effect seem to exist from graphical view. The analysis on developed and emerging jump types show that developed jumps cause an impact to emerging jumps more than the emerging jumps do to developed jumps. The developed jump self-excites itself more than an emerging jump and the developed jump has negative mean jump size on return whereas the emerging jump size on return is mixed. However, the parameters involving with the jump intensity are not statistically significant so stochastic jump intensity might not be needed. With the same reason, the analysis on jump effects asymmetries provides no evidence statistically.

In bivariate case, the model is able to show the statistical evidence on the jump clustering and jump contagion effects along with their asymmetries. The US market tends to react to most unexpected movements happening in the LA market while the LA market does not react much to the US market’s unexpected movements.

REFERENCES

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. J. (2014). Modeling financial contagion using mutually exciting jump processes. *National Bureau of Economic Research, Working Paper, (No. w15850)*.

Bacry, E., Delattre, S., Hoffman, M., & Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, *13*(1), 65-77.

Bacry, E., & Jean-Francois, M. (2013). Hawkes model for price and trades high-frequency dynamics. *arXiv preprint arXiv:1301.1135*.

Basu, S., & Dassios, A. (2002). A Cox process with log-normal intensity. *Insurance: mathematics and economics, 31*(2), 297-302.

Carlsson, J., Foo, M.-C., Lee, H.-H., & Shek, H. (2007). High Frequency Trade Prediction with Bivariate Hawkes Process. *Standford University*.

Chan, W. H., & Maheu, J. M. (2002). Conditional Jump Dynamics in Stock Market Returns. *Journal of Business & Economic Statistics*, *20*(3), 377-389.

Choe, G. H., & Lee, K. (2013). Conditional correlation in asset return and GARCH intensity model. *AStA Advances in Statistical Analysis*, 1-28.

Christoffersen, P., Kris, J., & Ornthanalai, C. (2012). Dynamic jump intensities and risk premiums: Evidence from S&P500 returns and options. *Journal of Financial Economics*, *106*(3), 447-472.

Couvreur, C. (1997). The EM Algorithm: A Guided Tour. *In Computer Intensive Methods in Control and Signal Processing*, 209-222.

Cox, D. R. (1955). Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society*, *17*(2), 129-164.

Dario, A. D., & Simonis, A. (2011). Properties of Doubly Stochastic Poisson Process with affine intensity. *arXiv preprint arXiv:1109.2884*.

Duncan, J., Randal, J., & Thomson, P. (2009). Fitting jump diffusion process using the EM algorithm. *Working Paper, Victoria University of Wellington*.

Eberly, D. (2003). Derivative approximation by finite differences. *Magic Software, Inc*.

Fang, H. (2000). Option Pricing Implications of a Stochastic Jump Rate. *Working Paper, University of Virginia*.

Fonseca, J. D., & Zaartour, R. (2014). Correlation and Lead-Lag Relationships in a Hawkes Microstructure. *Working Paper*.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83-90.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management science, 48*(8), 1086-1101.

Kou, S. G., & Wang, H. (2004). Option pricing under a double exponential jump diffusion model. *Management science, 50*(9), 1178-1192.

Li, G., & Zhang, C. (2013). Conditional Jump Intensity, Conditional Expected Jump Size, and Relative Stock Price Level. *Working Paper*.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics, 3*(1), 125-144.

Polson, N. G., & Scott, J. G. (2011). Explosive Volatility: A Model of Financial Contagion. *Working Paper*.

Roche, A. (2011). EM algorithm and variants: an informal tutorial. *arXiv preprint arXiv:1105.1476*.

Sepp, A. (2003). Fourier Transform for Option Pricing under Affine Jump-Diffusion: An Overview. *Unpublished Manuscript, available at www. hot. ee/seppar*.

APPENDIX

Appendix 1: Detailed Derivation

A1.1 Complete likelihood

The complete likelihood is given by:

$$L_c = P(\mathrm{X}, \mathrm{Z}, \Delta\mathrm{N}, \theta, \eta, \lambda \mid \Gamma)$$
$$= P(\lambda_{1,1}, \dots, \lambda_{K,1}) P(\mathrm{X}, \mathrm{Z}, \Delta\mathrm{N}, \theta, \eta, \lambda_{1,2}, \dots, \lambda_{K,T} \mid \lambda_{1,1}, \dots, \lambda_{K,1})$$

Let λ_t denote $\lambda_{1,t}, \dots, \lambda_{K,t}$
$\eta_{k,l,t}$ denote $\eta_{1,k,l,2}, \dots, \eta_{K,k,l,2}$
$\theta_{k,l,t}$ denote $\theta_{1,k,l,1}, \dots, \theta_{n,k,l,1}$
z_t denote $z_{1,t}, \dots z_{n,t}$
C_t denote all information up to time t then

$$= P(\lambda_1) \prod_{k=1}^{K} P(\Delta N_{k,1} | \lambda_{k,1}) P(\mathrm{X}, \mathrm{Z}, \theta, \eta, \Delta N_{1,2}, \dots, \Delta N_{K,T}, \lambda_{1,2}, \dots, \lambda_{1,T} | \Delta N_{k,1}, \lambda_1)$$

$$= P(\lambda_1) \prod_{k=1}^{K} P(\Delta N_{k,1} | \lambda_{k,1}) \prod_{j=1}^{\Delta N_{k,t}} P(\theta_{1,k,j,1}, \dots, \theta_{n,k,j,1}) \cdot P(z_{1,1}, \dots, z_{n,1})$$
$$\cdot P(\lambda_{1,2}, \dots, \lambda_{K,T}, \Delta N_{k,2}, \dots, \Delta N_{k,T}, \eta_{1,k,l,2}, \dots, \eta_{K,k,l,T}, z_{1,2}, \dots, z_{n,T}, \theta_{1,k,j,2}, \dots, \theta_{n,k,j,T} | C_1)$$

$$= P(\lambda_1) \prod_{k=1}^{K} P(\Delta N_{k,1} | \lambda_{k,1}) \prod_{j=1}^{\Delta N_{k,t}} P(\theta_{k,j,1}) \cdot P(z_1)$$

$$\cdot P(\lambda_2 | C_1) \prod_{k=1}^{K} P(\Delta N_{k,2} | \lambda_{k,2}, C_1) \prod_{l=1}^{\Delta N_{k,t}} P(\eta_{k,l,2} | C_1) \cdot \prod_{j=1}^{\Delta N_{k,t}} P(\theta_{k,j,2} | C_1) \cdot P(z_2 | C_1)$$
$$\vdots$$
$$\cdot P(\lambda_T | C_{T-1}) \prod_{k=1}^{K} P(\Delta N_{k,T} | \lambda_{k,T}, C_{T-1}) \prod_{l=1}^{\Delta N_{k,t}} P(\eta_{k,l,T} | C_{T-1}) \cdot \prod_{j=1}^{\Delta N_{k,t}} P(\theta_{k,j,T} | C_{T-1})$$
$$\cdot P(z_T | C_{T-1})$$

A1.2 Complete log-likelihood

The complete log-likelihood is given by:

$$\ln L_c = \sum_{t=1}^{T}\left\{\ln P(\lambda_t|C_{t-1}) + \sum_{k=1}^{K}\left[\ln P\left(\Delta N_{k,t}|\lambda_{k,t},C_{t-1}\right) + \sum_{l=1}^{\Delta N_{k,t}}\left\{\ln P\left(\eta_{k,l,t}|C_{t-1}\right) + \ln P\left(\theta_{k,l,t}|C_{t-1}\right)\right\}\right] + \ln P(z_t|C_{t-1})\right\}$$

where $[z_t]\sim Normal([\mu],[\Sigma])$

$[\theta_{k,l,t}]\sim Normal([\nu_k],[\Delta_k])$

$[\eta_{k,l,t}]\sim Normal([\kappa_k],[\Omega_k])$

$[u_t]\sim Normal([a],[B])$

$[\ln\lambda_1]\sim Normal([a_0],[B_0])$

$$\begin{aligned}\therefore \ln L_c = & -\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|B_0| - \frac{1}{2}(\ln\lambda_1 - a_0)'B_0^{-1}(\ln\lambda_1 - a_0) \\ & -\frac{K(T-1)}{2}\ln 2\pi - \frac{T-1}{2}\ln|B| \\ & -\frac{1}{2}\sum_{t=2}^{T}\left(\ln\lambda_t - a - c\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)' B^{-1}\left(\ln\lambda_t - a - c\ln\lambda_{j,t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right) \\ & + \sum_{t=1}^{T}\sum_{k=1}^{K}\left[-\lambda_{k,t} + \Delta N_{k,t}\ln\lambda_{k,t} - \ln\left(\Delta N_{k,t}!\right)\right] \\ & + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\left[-\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\left(\eta_{k,l,t-1} - \kappa_k\right)'\Omega^{-1}\left(\eta_{k,l,t-1} - \kappa_k\right)\right]\end{aligned}$$

$$+\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t}}\left[-\frac{K}{2}\ln 2\pi-\frac{1}{2}\ln|\Delta|-\frac{1}{2}\left(\theta_{k,l,t}-\nu_k\right)'\Delta^{-1}\left(\theta_{k,l,t}-\nu_k\right)\right]$$
$$-\frac{nT}{2}\ln 2\pi-\frac{T}{2}\ln|\Sigma|-\frac{1}{2}\sum_{t=1}^{T}(z_t-\mu)'\Sigma^{-1}(z_t-\mu)$$

A1.3 E-Step

The conditional expectation in E-Step is given by:

$$Q\left(\Gamma\middle|\Gamma^{(p)}\right)=\mathbb{E}[\ln L_C|\Gamma^{(p)},x_1,\dots,x_T]$$

Let X_t denotes $x_1,\dots,x_t$ while X denotes $x_1,\dots,x_T$

Note: most of the time, this research will omit $\Gamma^{(\mathrm{p})}$ *for better readability.*

$$Q\left(\Gamma\middle|\Gamma^{(p)}\right)=-\frac{K}{2}\ln 2\pi-\frac{1}{2}\ln|B_0|-\frac{1}{2}\mathbb{E}[(\ln\lambda_1-a_0)'B_0^{-1}(\ln\lambda_1-a_0)|\mathrm{X}]$$
$$-\frac{K(T-1)}{2}\ln 2\pi-\frac{T-1}{2}\ln|B|$$
$$-\mathbb{E}[\frac{1}{2}\sum_{t=2}^{T}\left(\ln\lambda_t-a-c\ln\lambda_{t-1}-\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)'B^{-1}\left(\ln\lambda_t-a-c\ln\lambda_{t-1}-\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)|\mathrm{X}]$$
$$+\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{K}-\lambda_{k,t}+\Delta N_{k,t}\ln\lambda_{k,t}-\ln\left(\Delta N_{k,t}!\right)\ \middle|\mathrm{X}\right]$$
$$+\mathbb{E}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\left\{\begin{matrix}-\frac{n}{2}\ln 2\pi-\frac{1}{2}\ln|\Omega|\\-\frac{1}{2}\left(\eta_{k,l,t-1}-\kappa_k\right)'\Omega^{-1}\left(\eta_{k,l,t-1}-\kappa_k\right)\end{matrix}\right\}\middle|\mathrm{X}\right]$$
$$+\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t}}-\frac{K}{2}\ln 2\pi-\frac{1}{2}\ln|\Delta|-\frac{1}{2}\left(\theta_{k,l,t}-\nu_k\right)'\Delta^{-1}\left(\theta_{k,l,t}-\nu_k\right)\middle|\mathrm{X}\right]$$
$$-\frac{nT}{2}\ln 2\pi-\frac{T}{2}\ln|\Sigma|-\frac{1}{2}\mathbb{E}[\sum_{t=1}^{T}(z_t-\mu)'\Sigma^{-1}(z_t-\mu)\,|\mathrm{X}]$$

A1.4 M-Step

The parameters' values that maximizes $Q(\Gamma|\Gamma^{(p)})$ in M-Step is given by:

Consider $(\hat{\mu},\hat{\Sigma})$;

$$\nabla_{\hat{\mu}}=0\ \rightarrow\ \nabla_{\hat{\mu}}\left(\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}\left[(z_t-\hat{\mu})'\hat{\Sigma}^{-1}(z_t-\hat{\mu})\middle|\mathrm{X}\right]\right)=0$$

$$\rightarrow \;\; \nabla_{\hat{\mu}}\left(\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}\left[\left(\hat{\Sigma}^{-\frac{1}{2}}z_t-\hat{\Sigma}^{-\frac{1}{2}}\mu\right)'\left(\hat{\Sigma}^{-\frac{1}{2}}z_t-\hat{\Sigma}^{-\frac{1}{2}}\mu\right)\Big|\mathrm{X}\right]\right)=0$$

from $\nabla_X(AX-b)'(AX-b)=2A'(AX-b)$

$$\rightarrow \;\; 2\hat{\Sigma}^{-1}\sum_{t=1}^{T}(\mathbb{E}[(z_t-\hat{\mu})|\mathrm{X}])=0$$

$$\rightarrow \;\; \hat{\mu}=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[z_t|\mathrm{X}]$$

$$\frac{\partial}{\partial\hat{\Sigma}}=0 \;\; \rightarrow \;\; \frac{\partial}{\partial\hat{\Sigma}}\left(-\frac{T}{2}\ln\left|\hat{\Sigma}\right| -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}\left[(z_t-\hat{\mu})'\hat{\Sigma}^{-1}(z_t-\hat{\mu})\middle|\mathrm{X}\right]\right)=0$$

from $\nabla_X \ln|X|=(X^{-1})'$ and $\nabla_X a'X^{-1}b=-(X^{-1})'ab'(X^{-1})'$

$$\rightarrow \;\; -T\left(\hat{\Sigma}^{-1}\right)'+(\hat{\Sigma}^{-1})'\mathbb{E}[\textstyle\sum_{t=1}^{T}(z_t-\hat{\mu})(z_t-\hat{\mu})'\,|\mathrm{X}](\hat{\Sigma}^{-1})'=0$$

$$\rightarrow \;\; \hat{\Sigma}=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[(z_t-\hat{\mu})(z_t-\hat{\mu})'|\mathrm{X}]$$

Consider $(\hat{\nu}_k,\hat{\Delta}_k)$;

$$\nabla_{\hat{\nu}_k}=0 \;\; \rightarrow \;\; \nabla_{\hat{\nu}_k}\left(-\frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\mathbb{E}\left[\textstyle\sum_{l=1}^{\Delta N_{k,t}}\left(\theta_{k,l,t}-\hat{\nu}_k\right)'\hat{\Delta}_k{}^{-1}\left(\theta_{k,l,t}-\hat{\nu}_k\right)\Big|\mathrm{X}\right]\right)=0$$

$$\rightarrow \;\; -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}\left[-2\Delta N_{k,t}\hat{\Delta}_k{}^{-1}\left(\theta_{k,1,t}-\hat{\nu}_k\right)\middle|\mathrm{X}\right]=0$$

$$\rightarrow \;\; \sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t}-\hat{\nu}_k\right)\middle|\mathrm{X}\right]=0$$

$$\rightarrow \;\; \sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t}\right)\middle|\mathrm{X}\right]=\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}(\hat{\nu}_k)\middle|\mathrm{X}\right]$$

$$\rightarrow \;\; \hat{\nu}_k \;=\; \frac{\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t}\right)\middle|\mathrm{X}\right]}{\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\middle|\mathrm{X}\right]}$$

$\nabla_{\hat{\Delta}_k}=0$

$$\rightarrow \;\; \nabla_{\hat{\Delta}_k}\left(\mathbb{E}\left[\textstyle\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t}}-\frac{1}{2}\ln\left|\hat{\Delta}_k\right|-\frac{1}{2}\left(\theta_{k,l,t}-\hat{\nu}_k\right)'\hat{\Delta}_k{}^{-1}\left(\theta_{k,l,t}-\hat{\nu}_k\right)\Big|\mathrm{X}\right]\right)=0$$

$$\rightarrow \;\; -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\middle|\mathrm{X}\right](\hat{\Delta}_k^{-1})'$$

$$+\frac{1}{2}\sum_{t=1}^{T}\left(\hat{\Delta}_k^{-1}\right)'\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t}-\hat{\nu}_k\right)\left(\theta_{k,1,t}-\hat{\nu}_k\right)'\Big|\mathrm{X}\right](\hat{\Delta}_k^{-1})'=0$$

$$\rightarrow \;\; \hat{\Delta}_k=\frac{\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\left(\theta_{k,1,t}-\hat{\nu}_k\right)\left(\theta_{k,1,t}-\hat{\nu}_k\right)'\Big|\mathrm{X}\right]}{\sum_{t=1}^{T}\mathbb{E}\left[\Delta N_{k,t}\middle|\mathrm{X}\right]}$$

Consider $(\hat{\kappa}_k, \hat{\Omega}_k)$;

$$\nabla_{\hat{\kappa}_k} = 0 \;\rightarrow\; \nabla_{\hat{\kappa}_k}\left(-\frac{1}{2}\sum_{t=2}^{T}\sum_{k=1}^{K}\mathbb{E}\left[\sum_{l=1}^{\Delta N_{k,t-1}}\left(\eta_{k,l,t}-\hat{\kappa}_k\right)'\hat{\Omega}_k^{\;-1}\left(\eta_{k,l,t}-\hat{\kappa}_k\right)\middle|\mathrm{X}\right]\right) = 0$$

$$\rightarrow\; -\frac{1}{2}\sum_{t=2}^{T}\mathbb{E}\left[-2\Delta N_{k,t-1}\hat{\Omega}_k^{\;-1}\left(\eta_{k,1,t-1}-\hat{\kappa}_k\right)\middle|\mathrm{X}\right] = 0$$

$$\rightarrow\; \sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1}-\hat{\kappa}_k\right)\middle|\mathrm{X}\right] = 0$$

$$\rightarrow\; \sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1}\right)\middle|\mathrm{X}\right] = \sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}(\hat{\kappa}_k)\middle|\mathrm{X}\right]$$

$$\rightarrow\; \hat{\kappa}_k = \frac{\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1}\right)\middle|\mathrm{X}\right]}{\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\middle|\mathrm{X}\right]}$$

$\nabla_{\hat{\Omega}_k} = 0$

$$\rightarrow\; \nabla_{\hat{\Omega}_k}\left(\mathbb{E}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}} -\frac{1}{2}\ln\left|\hat{\Omega}_k\right| - \frac{1}{2}\left(\eta_{k,l,t-1}-\hat{\kappa}_k\right)'\hat{\Omega}_k^{\;-1}\left(\eta_{k,l,t-1}-\hat{\kappa}_k\right)\middle|\mathrm{X}\right]\right) = 0$$

$$\rightarrow\; -\frac{1}{2}\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\middle|\mathrm{X}\right](\hat{\Omega}_k^{-1})' + \frac{1}{2}\sum_{t=2}^{T}(\hat{\Omega}_k^{-1})'\,\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1}-\hat{\kappa}_k\right)\left(\eta_{k,1,t-1}-\hat{\kappa}_k\right)'\middle|\mathrm{X}\right](\hat{\Omega}_k^{-1})' = 0$$

$$\rightarrow\; \hat{\Omega}_k = \frac{\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1}-\hat{\kappa}_k\right)\left(\eta_{k,1,t-1}-\hat{\kappa}_k\right)'\middle|\mathrm{X}\right]}{\sum_{t=2}^{T}\mathbb{E}\left[\Delta N_{k,t-1}\middle|\mathrm{X}\right]}$$

Consider $(\hat{a}_0, \hat{B}_0, \hat{a}, \hat{B})$;

$$\nabla_{\hat{a}_0} = 0 \;\rightarrow\; \nabla_{\hat{a}_0}\left(\mathbb{E}\left[(\ln\lambda_1-\hat{a}_0)'\hat{B}_0^{\;-1}(\ln\lambda_1-\hat{a}_0)\middle|\mathrm{X}\right]\right) = 0$$

$$\rightarrow\; \mathbb{E}[(\ln\lambda_1-\hat{a}_0)|\mathrm{X}] = 0$$

$$\rightarrow\; \hat{a}_0 = \mathbb{E}[\ln\lambda_1|\mathrm{X}]$$

$$\nabla_{\hat{B}_0} = 0 \;\rightarrow\; \nabla_{\hat{B}_0}\left(-\frac{1}{2}\ln\left|\hat{B}_0\right| - \frac{1}{2}\mathbb{E}\left[(\ln\lambda_1-\hat{a}_0)'\hat{B}_0^{\;-1}(\ln\lambda_1-\hat{a}_0)\middle|\mathrm{X}\right]\right) = 0$$

$$\rightarrow\; -\left(\hat{B}_0^{\;-1}\right)' + (\hat{B}_0^{\;-1})'\mathbb{E}[(\ln\lambda_1-\hat{a}_0)(\ln\lambda_1-\hat{a}_0)'|\mathrm{X}](\hat{B}_0^{\;-1})' = 0$$

$$\rightarrow\; \hat{B}_0 = \mathbb{E}[(\ln\lambda_1-\hat{a}_0)(\ln\lambda_1-\hat{a}_0)'|\mathrm{X}]$$

$$\nabla_{\hat{a}}= 0 \;\rightarrow\; \nabla_{\hat{a}}(-\frac{1}{2}\mathbb{E}[\sum_{t=2}^{T}\left(\ln\lambda_t - \hat{a} - c\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)'\hat{B}^{-1}\Big(\ln\lambda_t$$

$$- \hat{a} - \ln\lambda_{j,t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\Big)|\mathrm{X}]) \;= 0$$

$$\rightarrow\; \nabla_{\hat{a}}\left(-\frac{1}{2}\mathbb{E}\left[\sum_{t=2}^{T}(u_t-\hat{a})'\hat{B}^{-1}(u_t-\hat{a})\,\middle|\mathrm{X}\right]\right) = 0$$

$$\rightarrow\; \sum_{t=2}^{T}\mathbb{E}[(u_t-\hat{a})|\mathrm{X}] = 0$$

$$\rightarrow\; \hat{a} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}[u_t|\mathrm{X}]$$

$$\nabla_{\hat{B}}= 0 \;\rightarrow\; \nabla_{\hat{B}}\{-\frac{T-1}{2}\ln\left|\hat{B}\right|$$

$$-\mathbb{E}[\frac{1}{2}\sum_{t=2}^{T}\left(\ln\lambda_t - \hat{a} - \hat{c}\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\right)'\hat{B}^{-1}\Big(\ln\lambda_t$$

$$- \hat{a} - \hat{c}\ln\lambda_{t-1} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\Big)|\mathrm{X}]\} = 0$$

$$\rightarrow\; \nabla_{\hat{B}}\left\{-\frac{T-1}{2}\ln\left|\hat{B}\right| - \mathbb{E}\left[\frac{1}{2}\sum_{t=2}^{T}(u_t-\hat{a})'\hat{B}^{-1}(u_t-\hat{a})\,\middle|\mathrm{X}\right]\right\} = 0$$

$$\rightarrow\; -(T-1)\left(\hat{B}^{-1}\right)' + (\hat{B}^{-1})'\mathbb{E}[\sum_{t=2}^{T}(u_t-\hat{a})(u_t-\hat{a})'\,|\mathrm{X}](\hat{B}^{-1})' = 0$$

$$\rightarrow\; \hat{B} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}[(u_t-\hat{a})(u_t-\hat{a})'|\mathrm{X}]$$

$$\nabla_{\hat{c}}= 0 \;\rightarrow\; \nabla_{\hat{c}}\{-\frac{T-1}{2}\ln\left|\hat{B}\right|$$

$$-\mathbb{E}[\frac{1}{2}\sum_{t=2}^{T}\Big(\ln\lambda_t - \hat{a} - diag(\ln\lambda_{t-1})\hat{c}$$

$$-\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\Big)'\hat{B}^{-1}\Big(\ln\lambda_t - \hat{a} - diag(\ln\lambda_{t-1})\hat{c}$$

$$-\sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}}\eta_{k,l,t-1}\Big)|\mathrm{X}]\} = 0$$

$$\rightarrow \; 2\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1} diag(\ln\lambda_{t-1}) \,\middle|\, \mathrm{X}\right]\hat{c} - 2\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}(\ln\lambda_t - \hat{a} - \sum_{k=1}^{K}\sum_{l=1}^{\Delta N_{k,t-1}} \eta_{k,l,t-1}) \,\middle|\, \mathrm{X}\right] = 0$$

$$\rightarrow \; \mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1} diag(\ln\lambda_{t-1}) \,\middle|\, \mathrm{X}\right]\hat{c} - \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a} + diag(\ln\lambda_{t-1})c^{(p)}) \,\middle|\, \mathrm{X}\right] = 0$$

$$\rightarrow \; \hat{c} = \mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1} diag(\ln\lambda_{t-1}) \,\middle|\, \mathrm{X}\right]^{-1} \times \mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a} + diag(\ln\lambda_{t-1})c^{(p)}) \,\middle|\, \mathrm{X}\right]$$

$$\rightarrow \; \hat{c} = c^{(p)} + \left(\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1} diag(\ln\lambda_{t-1}) \,\middle|\, \mathrm{X}\right]^{-1} \times \mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a}) \,\middle|\, \mathrm{X}\right]\right)$$

A1.5 Further Derivation on Parameters Estimation for E-Step

Consider Forward probability: $\alpha(\lambda_t, \Delta N_t, x_t | X_{t-1})$

$$\alpha(\lambda_1, \Delta N_1, x_1) = P(\lambda_1, \Delta N_1, x_1)$$

$$= P(\lambda_1)\prod_{k=1}^{K}\{P(\Delta N_{k,1}|\lambda_1)\}P(x_1|\Delta N_1, \lambda_1)$$

$$= P(\lambda_1)\prod_{k=1}^{K}\{P(\Delta N_{k,1}|\lambda_1)\}P(x_1|\Delta N_1)$$

$$= Normal(\ln\lambda_1; a_0, B_0)\cdot\prod_{k=1}^{K}\{Poisson(\Delta N_{k,1}; \lambda_{k,1})\} \cdot Normal\left(x_1; \mu + \sum_{k=1}^{K}\Delta N_{k,1}\nu_k, \Sigma + \sum_{k=1}^{K}\Delta N_{k,1}\Delta_k\right)$$

$$= \frac{1}{\sqrt{(2\pi)^K|B_0|}} e^{-\frac{1}{2}(\ln\lambda_1 - a_0)' B_0^{-1}(\ln\lambda_1 - a_0)} \cdot \prod_{k=1}^{K}\frac{e^{-\lambda_{k,1}}\lambda_{k,1}^{\Delta N_{k,1}}}{\Delta N_{k,1}!} \cdot \frac{1}{\sqrt{(2\pi)^n|\Sigma + \sum_{k=1}^{K}\Delta N_{k,1}\Delta_k|}} e^{-\frac{1}{2}\left(x_1 - \mu - \sum_{k=1}^{K}\Delta N_{k,1}\nu_k\right)'\left(\Sigma + \sum_{k=1}^{K}\Delta N_{k,1}\Delta_k\right)^{-1}\left(x_1 - \mu - \sum_{k=1}^{K}\Delta N_{k,1}\nu_k\right)}$$

Let $\alpha_t(\mathrm{a}, \mathrm{b}, \mathrm{c})$ *denotes* $\alpha_t(\lambda_1 = \mathrm{a}, \Delta N_1 = \mathrm{b}, x_1 = \mathrm{c})$ *where t is the time*

$$\alpha_1(\lambda, \mathrm{n}) = \alpha(\lambda_1, \Delta N_1, x_1)$$

$$\alpha_2(\lambda, \mathrm{n}) = \alpha(\lambda, \mathrm{n}, x_2)$$

$$= P(\lambda_2 = \lambda, \Delta N_2 = \mathrm{n}, x_2 | x_1)$$

$$= \sum_{\lambda_1, \mathrm{m}} P(\lambda_2 = \lambda, \Delta N_2 = \mathrm{n}, x_2 | \lambda_1, \Delta N_1 = \mathrm{m}, x_1)\frac{\alpha_1(\lambda_1, \mathrm{m})}{P(x_1)}$$

$$= \sum_{\lambda_1, \mathrm{m}} P(\Delta N_2 = \mathrm{n}, x_2 | \lambda_2 = \lambda, \lambda_1, \Delta N_1 = \mathrm{m}, x_1)P(\lambda_2 = \lambda | \lambda_1, \Delta N_1 = \mathrm{m}, x_1)\frac{\alpha_1(\lambda_1, \mathrm{m})}{P(x_1)}$$

$$
\begin{aligned}
&= \sum_{\lambda_1,\mathrm{m}} P(x_2|\Delta N_2 = \mathrm{n}, \lambda_2 = \lambda, \lambda_1, \Delta N_1 \\
&\qquad = \mathrm{m}, x_1)\, P(\Delta N_2 = \mathrm{n}|\lambda_2 = \lambda, \lambda_1, \Delta N_1 = \mathrm{m}, x_1) \\
&\qquad P(\lambda_2 = \lambda|\lambda_1, \Delta N_1 = \mathrm{m}, x_1)\, \frac{\alpha_1(\lambda_1, \mathrm{m})}{P(x_1)} \\
&= \sum_{\lambda_1,\mathrm{m}} P(x_2|\Delta N_2 = \mathrm{n}) \prod_{k=1}^{K} \left\{ P\left(\Delta N_{k,2} = \mathrm{n}_k \middle| \lambda_2 = \lambda\right) \right\} P(\lambda_2 = \lambda|\lambda_1, \Delta N_1 \\
&\qquad = \mathrm{m})\, \frac{\alpha_1(\lambda_1, \mathrm{m})}{P(x_1)} \\
&= \sum_{\lambda_1,\mathrm{m}} Normal\left(x_2; \mu + \sum_{k=1}^{K} \mathrm{n}_k \nu_k\, , \Sigma \right. \\
&\qquad \left. + \sum_{k=1}^{K} \mathrm{n}_k \Delta_k \right) \prod_{k=1}^{K} \{ Poisson(\mathrm{n}_k; \lambda_k) \}\, Normal(\ln \lambda\, ; a + \ln \lambda_1 \\
&\qquad + \sum_{k=1}^{K} \mathrm{m}_k \kappa_k\, , B + \sum_{k=1}^{K} \mathrm{m}_k \Omega_k)\, \frac{\alpha_1(\lambda_1, \mathrm{m})}{P(x_1)}
\end{aligned}
$$

$$
\begin{aligned}
\alpha_t(\lambda, \mathrm{n}) &= P(\lambda_t = \lambda, \Delta N_\mathrm{t} = \mathrm{n}, x_t|\mathrm{X}_{\mathrm{t}-1}) \\
&= \sum_{\lambda_{t-1},\mathrm{m}} P(\lambda_t = \lambda, \Delta N_\mathrm{t} = \mathrm{n}, x_t|\lambda_{t-1}, \Delta N_{t-1} \\
&\qquad = \mathrm{m}, \mathrm{X}_{t-1})\, P(\lambda_{t-1}, \Delta N_{t-1} = \mathrm{m}|\mathrm{X}_{\mathrm{t}-1}) \\
&= \sum_{\lambda_{t-1},\mathrm{m}} P(x_\mathrm{t}|\Delta N_t = \mathrm{n}, \lambda_t = \lambda, \lambda_{t-1}, \Delta N_{t-1} = \mathrm{m}, \mathrm{X}_{\mathrm{t}-1}) \\
&\qquad P(\Delta N_t = \mathrm{n}|\lambda_t = \lambda, \lambda_{t-1}, \Delta N_{t-1} = \mathrm{m}, \mathrm{X}_{t-1}) \\
&\qquad P(\lambda_t = \lambda|\lambda_{t-1}, \Delta N_{t-1} = \mathrm{m}, \mathrm{X}_{\mathrm{t}-1})\, \frac{\alpha_{t-1}(\lambda_{t-1}, \mathrm{m})}{P(x_{t-1}|\mathrm{X}_{\mathrm{t}-2})} \\
&= \sum_{\lambda_{t-1},\mathrm{m}} P(x_\mathrm{t}|\Delta N_t = \mathrm{n}) \prod_{k=1}^{K} \left\{ P\left(\Delta N_{k,t} = \mathrm{n}_k \middle| \lambda_{k,t} = \lambda_k\right) \right\} \\
&\qquad \cdot P(\lambda_t = \lambda|\lambda_{t-1}, \Delta N_{t-1} = \mathrm{m}) \cdot \frac{\alpha_{t-1}(\lambda_{t-1}, \mathrm{m})}{P(x_{t-1}|\mathrm{X}_{\mathrm{t}-2})}
\end{aligned}
$$

$$
\begin{aligned}
\alpha_t(\lambda, \mathrm{n}) &= \sum_{\lambda_{t-1},\mathrm{m}} Normal(x_t; \mu + \sum_{k=1}^{K} \mathrm{n}_k \nu_k\, , \Sigma + \sum_{k=1}^{K} \mathrm{n}_k \Delta_k) \cdot \prod_{k=1}^{K} Poisson(\mathrm{n}_k; \lambda_k) \\
&\qquad \cdot Normal\left(\ln \lambda\, ; a + \ln \lambda_{t-1} + \sum_{k=1}^{K} \mathrm{m}_k \kappa_k\, , B + \sum_{k=1}^{K} \mathrm{m}_k \Omega_k \right) \\
&\qquad \cdot \frac{\alpha_{t-1}(\lambda_{t-1}, \mathrm{m})}{P(x_{t-1}|\mathrm{X}_{\mathrm{t}-2})}
\end{aligned}
$$

Consider Backward probability: $\gamma(\lambda_t, \Delta N_t|\mathrm{X})$

$$\gamma_T(\lambda, \mathrm{n}) = P(\lambda_T = \lambda, \Delta N_T = \mathrm{n}|\mathrm{X}) = \frac{P(\lambda_T = \lambda, \Delta N_T = \mathrm{n}, x_T|\mathrm{X_{t-1}})}{P(x_T|\mathrm{X_{T-1}})} = \frac{\alpha_T(\lambda, \mathrm{n})}{P(x_T|\mathrm{X_{T-1}})} = \frac{\alpha_T(\lambda, \mathrm{n})}{\sum_{\mathrm{r,s}} \alpha_T(\mathrm{r}, \mathrm{s})}$$

$$\gamma_t(\lambda, \mathrm{n}) = P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}|\mathrm{X}) = \sum_{\mathrm{r}}^{\infty} P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}, \lambda_{t+1} = \mathrm{r}|\mathrm{X}) = \sum_{\mathrm{r}}^{\infty} \zeta_t(\lambda, \mathrm{n}, \mathrm{r})$$

Consider 2-Step probability: $\zeta(\lambda_t, \Delta N_t, \lambda_{t+1}|\mathrm{X})$

$$\zeta_t(\lambda, \mathrm{n}, \mathrm{r}) = P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}, \lambda_{t+1} = \mathrm{r}|\mathrm{X}) = \sum_{\mathrm{m}}^{\infty} P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}, \lambda_{t+1} = \mathrm{r}, \Delta N_{t+1} = \mathrm{m}|\mathrm{X})$$

$$= \sum_{\mathrm{m}}^{\infty} P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}|\lambda_{t+1} = \mathrm{r}, \Delta N_{t+1} = \mathrm{m}, \mathrm{X_t})\gamma_{\mathrm{t+1}}(\mathrm{r}, \mathrm{m})$$

$$= \sum_{\mathrm{m}}^{\infty} \frac{P(\lambda_{t+1} = \mathrm{r}, \Delta N_{t+1} = \mathrm{m}|\lambda_t = \lambda, \Delta N_t = \mathrm{n}, \mathrm{X_t})P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}|\mathrm{X_t})}{P(\lambda_{t+1} = \mathrm{r}, \Delta N_{t+1} = \mathrm{m}|\mathrm{X_t})}\gamma_{\mathrm{t+1}}(\mathrm{r}, \mathrm{m})$$

$$= \sum_{\mathrm{m}}^{\infty} \left\{ \frac{\begin{array}{c} Poisson(\mathrm{m}; \mathrm{r}) \\ \cdot Normal(\ln \mathrm{r}\,; a + \ln \lambda + \sum_{k=1}^{K} \mathrm{n}_k \kappa_k\,, B + \sum_{k=1}^{K} n_k \Omega_k) \\ \cdot \frac{\alpha_t(\lambda, \mathrm{n})}{P(x_t|\mathrm{X_{t-1}})} \end{array}}{\begin{array}{c} \sum_{\mathrm{s,p}}[Poisson(\mathrm{m}; \mathrm{r}) \\ \cdot Normal(\ln \mathrm{r}\,; a + \ln \mathrm{s} + \sum_{k=1}^{K} \mathrm{p}_k \kappa_k\,, B + \sum_{k=1}^{K} \mathrm{p}_k \Omega_k) \\ \cdot \frac{\alpha_t(\mathrm{s}, \mathrm{p})}{P(x_t|\mathrm{X_{t-1}})}] \end{array}} \cdot \gamma_{\mathrm{t+1}}(\mathrm{r}, \mathrm{m}) \right\}$$

$$= \left(\sum_{\mathrm{m}}^{\infty} \gamma_{\mathrm{t+1}}(\mathrm{r}, \mathrm{m}) \right) \cdot \left(\frac{Normal(\ln \mathrm{r}\,; a + \ln \lambda + \sum_{k=1}^{K} \mathrm{n}_k \kappa_k\,, B + \sum_{k=1}^{K} n_k \Omega_k)\alpha_t(\lambda, \mathrm{n})}{\sum_{\mathrm{s,p}} Normal(\ln \mathrm{r}\,; a + \ln \mathrm{s} + \sum_{k=1}^{K} \mathrm{p}_k \kappa_k\,, B + \sum_{k=1}^{K} \mathrm{p}_k \Omega_k)\alpha_t(\mathrm{s}, \mathrm{p})} \right)$$

Consider $P(\Delta N_t = \mathrm{n}|\mathrm{X})$;

$$P(\Delta N_t = \mathrm{n}|\mathrm{X}) = \sum_{\lambda} P(\lambda_t = \lambda, \Delta N_t = \mathrm{n}|\mathrm{X}) = \sum_{\lambda} \gamma_t(\lambda, \mathrm{n})$$

if $\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \sim Normal$ with mean $\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and variance matrix $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

then $\mathbb{E}[S_1|S_2 = a] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$ (A1)

and $Var(S_1|S_2 = a] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (A2)

Consider $\mathbb{E}[\Delta N_{k,t}\mathrm{X}]$;

$$\mathbb{E}[\Delta N_{k,t}|\mathrm{X}] = \sum_{\mathrm{n}=0}^{\infty} \mathbb{E}[\Delta N_{k,t}|\Delta N_{k,t} = \mathrm{n}_k, \mathrm{X}]P(\Delta N_{k,t} = \mathrm{n}_k|\mathrm{X})$$
$$= \sum_{\mathrm{n}=1}^{\infty} \mathrm{n}_k \cdot P(\Delta N_{k,t} = \mathrm{n}_k|\mathrm{X})$$

Consider $\mathbb{E}[z_t|\mathrm{X}]$;

$$\mathbb{E}[z_t|\mathrm{X}] = \sum_{\mathrm{n}=0}^{\infty} \mathbb{E}[z_t|\Delta N_t = \mathrm{n}, \mathrm{X}]P(\Delta N_t = \mathrm{n}|\mathrm{X})$$
$$= \sum_{\mathrm{n}=0}^{\infty} \mathbb{E}[z_t|\Delta N_t = \mathrm{n}, x_t]P(\Delta N_t = \mathrm{n}|\mathrm{X})$$

From (A1) and (A2),

$$\mathbb{E}[z_t|x_t, \Delta N_t = \mathrm{n}] = \hat{\mu} + \hat{\Sigma}(\hat{\Sigma} + \mathrm{n}\hat{\Delta})^{-1}(x_t - (\hat{\mu} + \mathrm{n}\hat{\nu}))$$
$$Var[z_t|x_t, \Delta N_t = \mathrm{n}] = \hat{\Sigma} - \hat{\Sigma}(\hat{\Sigma} + \mathrm{n}\hat{\Delta})^{-1}\hat{\Sigma}$$

Consider $\mathbb{E}[(z_t - \hat{\mu})(z_t - \hat{\mu})'|\mathrm{X}]$;

$$\mathbb{E}[(z_t - \hat{\mu})(z_t - \hat{\mu})'|\mathrm{X}]$$
$$= \sum_{\mathrm{n}=0}^{\infty} \mathbb{E}[(z_t - \hat{\mu})(z_t - \hat{\mu})'|\mathrm{X}, \Delta N_t = \mathrm{n}]\, P(\Delta N_t = \mathrm{n}|\mathrm{X})$$
$$= \sum_{\mathrm{n}=0}^{\infty} \mathbb{E}[(z_t - \hat{\mu})(z_t - \hat{\mu})'|x_t, \Delta N_t = \mathrm{n}]\, P(\Delta N_t = \mathrm{n}|\mathrm{X})$$

where $\mathbb{E}[(z_t - \hat{\mu})(z_t - \hat{\mu})'|x_t, \Delta N_t = \mathrm{n}]$

$$= \mathbb{E}\left[\begin{matrix} Var(z_t|x_t, \Delta N_t = \mathrm{n}) \\ +(\mathbb{E}[z_t|x_t, \Delta N_t = \mathrm{n}] - \hat{\mu})(\mathbb{E}[z_t|x_t, \Delta N_t = \mathrm{n}] - \hat{\mu})' \end{matrix} \middle| x_t, \Delta N_t = \mathrm{n}\right]$$
$$= \hat{\Sigma} - \hat{\Sigma}(\hat{\Sigma} + \mathrm{n}\hat{\Delta})^{-1}\hat{\Sigma}$$
$$+ \left(\hat{\Sigma}(\hat{\Sigma} + \mathrm{n}\hat{\Delta})^{-1}(x_t - (\hat{\mu} + \mathrm{n}\hat{\nu}))\right)\left(\hat{\Sigma}(\hat{\Sigma} + \mathrm{n}\hat{\Delta})^{-1}(x_t - (\hat{\mu} + \mathrm{n}\hat{\nu}))\right)$$

Consider $\mathbb{E}[\,\Delta N_{k,t}(\theta_{k,1,t})|\mathrm{X}]$;

$$\mathbb{E}[\,\Delta N_{k,t}(\theta_{k,1,t})|\mathrm{X}] = \sum_{\mathrm{n}_k} \Delta N_{k,t} \cdot \mathbb{E}[\,\theta_{k,1,t}|\mathrm{X}, \Delta N_{k,t} = \mathrm{n}_k] \cdot \mathbb{P}(\Delta N_{k,t} = \mathrm{n}_k|\mathrm{X})$$

$$= \sum_{\mathrm{n}_k} \mathrm{n}_k \cdot \mathbb{E}[\,\theta_{k,1,t}|x_t, \Delta N_{k,t} = \mathrm{n}_k] \cdot \mathbb{P}(\Delta N_{k,t} = \mathrm{n}_k|\mathrm{X})$$

where $\mathbb{E}[\,\theta_{k,1,t}|x_t, \Delta N_{k,t} = \mathrm{n}_k] = \hat{v}_k + \hat{\Delta}_k(\hat{\Sigma} + \mathrm{n}_k\hat{\Delta}_k)^{-1}(x_t - (\hat{\mu} + \mathrm{n}_k\hat{v}_k\,))$

Consider $\mathbb{E}\left[\Delta N_{k,t}(\theta_{k,1,t} - \hat{v}_k)(\theta_{k,1,t} - \hat{v}_k)'\middle|\,\mathrm{X}\right]$;

$$\mathbb{E}\left[\Delta N_{k,t}(\theta_{k,1,t} - \hat{v}_k)(\theta_{k,1,t} - \hat{v}_k)'\middle|\,\mathrm{X}\right]$$

$$= \sum_{\mathrm{n}_k} \Delta N_{k,t} \cdot \mathbb{E}\left[(\theta_{k,1,t} - \hat{v}_k)(\theta_{k,1,t} - \hat{v}_k)'\middle|\,\mathrm{X}, \Delta N_{k,t} = \mathrm{n}_k\right] \cdot \mathbb{P}(\Delta N_{k,t} = \mathrm{n}_k|\mathrm{X})$$

where $\mathbb{E}\left[(\theta_{k,1,t} - \hat{v}_k)(\theta_{k,1,t} - \hat{v}_k)'\middle|\,\mathrm{X}, \Delta N_{k,t} = \mathrm{n}_k\right]$

$$= \mathbb{E}\Big[Var(\theta_{k,1,t}|x_t, \mathrm{n}_k) + (\mathbb{E}[\theta_{k,1,t}|x_t, \mathrm{n}_k] - \hat{v}_k)(\mathbb{E}[\theta_{k,1,t}|x_t, \mathrm{n}_k] - \hat{v}_k)'|\mathrm{X}_t, \Delta N_{k,t} = n_k\Big]$$

$$= \hat{\Delta}_k - \hat{\Delta}_k(\hat{\Sigma} + \mathrm{n}_k\hat{\Delta}_k)^{-1}\hat{\Delta}_k + \Big(\hat{\Delta}_k(\hat{\Sigma} + \mathrm{n}_k\hat{\Delta}_k)^{-1}(x_t - (\hat{\mu} + \mathrm{n}_k\hat{v}_k\,))\Big)\Big(\hat{\Delta}_k(\hat{\Sigma} + \mathrm{n}_k\hat{\Delta}_k)^{-1}(x_t - (\hat{\mu} + \mathrm{n}_k\hat{v}_k\,))\Big)'$$

Consider $\mathbb{E}[\Delta N_{k,t-1}(\eta_{k,1,t})|\mathrm{X}]$;

$$\mathbb{E}[\Delta N_{k,t-1}(\eta_{k,1,t})|\mathrm{X}]$$

$$= \sum_{\mathrm{n}_k=0}^{\infty} \mathbb{E}[\Delta N_{k,t-1}(\eta_{k,1,t})|\mathrm{X}, \Delta N_{k,t-1} = \mathrm{n}_k] P(\Delta N_{k,t-1} = \mathrm{n}_k|\mathrm{X})$$

$$= \sum_{\mathrm{n}_k=1}^{\infty} \mathrm{n}_k \cdot \mathbb{E}[\,\eta_{k,1,t}|\mathrm{X}, \Delta N_{k,t-1} = \mathrm{n}_k] P(\Delta N_{k,t-1} = \mathrm{n}_k|\mathrm{X})$$

$$= \sum_{\mathrm{n}_k=1}^{\infty} \left[\begin{array}{c} \mathrm{n}_k \cdot \sum_{\mathrm{r}_k,\mathrm{s}_k} \mathbb{E}[\,\eta_{k,1,t}|\mathrm{X}, \Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k] \\ \cdot P(\lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k|\mathrm{X}, \Delta N_{k,t-1} = \mathrm{n}_k) P(\Delta N_{k,t-1} = \mathrm{n}_k|\mathrm{X}) \end{array}\right]$$

$$= \sum_{\mathrm{n}_k=1}^{\infty} \left[\begin{array}{c} \mathrm{n}_k \cdot \sum_{\mathrm{r}_k,\mathrm{s}_k} \mathbb{E}[\,\eta_{k,1,t}|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k] \\ \cdot \dfrac{P(\lambda_{k,t-1} = \mathrm{r}_k, \Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t} = \mathrm{s}_k|\mathrm{X})}{P(\Delta N_{k,t-1} = \mathrm{n}_k|\mathrm{X})} P(\Delta N_{k,t-1} = \mathrm{n}_k|\mathrm{X}) \end{array}\right]$$

$$= \sum_{\mathrm{n}_k=1}^{\infty} \left[\mathrm{n}_k \cdot \sum_{\mathrm{r}_k,\mathrm{s}_k} \mathbb{E}[\,\eta_{k,1,t}|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k] \cdot \zeta_{t-1}(\mathrm{r}_k, \mathrm{n}_k, \mathrm{s}_k)\right]$$

$$= \sum_{\mathrm{n}_k=1}^{\infty}\left[\begin{matrix}\mathrm{n}_k \cdot \zeta_{t-1}(\mathrm{r}_k, \mathrm{n}_k, \mathrm{s}_k) \\ \cdot \sum_{\mathrm{r}_k,\mathrm{s}_k}\left(\hat{\kappa}_k + \hat{\Omega}_k\left(\hat{B} + \mathrm{n}_k\hat{\Omega}_k\right)^{-1}(\ln \mathrm{s}_k - \mathrm{c} \cdot \ln \mathrm{r}_k - \hat{a} - \mathrm{n}_k \cdot \hat{\kappa}_k)\right)\end{matrix}\right]$$

where the last eqution comes from (A1).

Consider $\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,\mathrm{X}\right]$;

$$\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,\mathrm{X}\right]$$

$$= \sum_{\mathrm{n}_k=0}^{\infty}\mathbb{E}\left[\Delta N_{k,t-1}\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,\mathrm{X}, \Delta N_{k,t-1} = \mathrm{n}_k\right]P\left(\Delta N_{k,t-1} = \mathrm{n}_k\middle|\mathrm{X}\right)$$

$$= \sum_{\mathrm{n}_k=1}^{\infty}\mathrm{n}_k \cdot \mathbb{E}\left[\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,x_t, \Delta N_{k,t-1} = \mathrm{n}_k\right]P(\Delta N_{k,t-1} = \mathrm{n}_k|\mathrm{X})$$

$$= \sum_{\mathrm{n}_k=1}^{\infty}\left[\mathrm{n}_k \cdot \sum_{\mathrm{r}_k,\mathrm{s}_k}\mathbb{E}\left[\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right] \cdot P\left(\lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\middle|\mathrm{X}, \Delta N_{k,t-1} = \mathrm{n}_k\right)P\left(\Delta N_{k,t-1} = \mathrm{n}_k\middle|\mathrm{X}\right)\right]$$

$$= \sum_{\mathrm{n}_k=1}^{\infty}\left[\mathrm{n}_k \cdot \sum_{\mathrm{r}_k,\mathrm{s}_k}\mathbb{E}\left[\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,\mathrm{n}_k, \mathrm{r}_k, \mathrm{s}_k\right] \cdot \zeta_{t-1}(\mathrm{r}_k, \mathrm{n}_k, \mathrm{s}_k)\right]$$

where $\mathbb{E}\left[\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)\left(\eta_{k,1,t-1} - \hat{\kappa}_k\right)'\middle|\,\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right] =$

$$\mathbb{E}\left[\begin{matrix} Var\left(\eta_{k,1,t-1}\middle|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right) \\ + \left(\mathbb{E}\left[\eta_{k,1,t-1}\middle|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right] - \hat{\kappa}_k\right) \\ \cdot \left(\mathbb{E}\left[\eta_{k,1,t-1}\middle|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right] - \hat{\kappa}_k\right)'\end{matrix}\middle|\, x_t, \Delta N_{k,t-1} = \mathrm{n}_k\right]$$

$$= \hat{\Omega}_k - \hat{\Omega}_k\left(\hat{B} + \mathrm{n}_k\hat{\Omega}_k\right)^{-1}\hat{\Omega}_k + \left(\mathbb{E}\left[\eta_{k,1,t-1}\middle|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right] - \hat{\kappa}_k\right)\left(\mathbb{E}\left[\eta_{k,1,t-1}\middle|\Delta N_{k,t-1} = \mathrm{n}_k, \lambda_{k,t-1} = \mathrm{r}_k, \lambda_{k,t} = \mathrm{s}_k\right] - \hat{\kappa}_k\right)'$$

Consider $\mathbb{E}[\ln \lambda_1|\mathrm{X}]$;

$$\mathbb{E}[\ln \lambda_1|\mathrm{X}] = \sum_{\mathrm{r,n,s}}\mathbb{E}[\ln \lambda_1|\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}, \mathrm{X}]P(\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = s|\mathrm{X})$$

$$= \sum_{\mathrm{r,n,s}}\mathbb{E}[\ln \lambda_1|\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}] \cdot \zeta_1(\mathrm{r}, \mathrm{n}, \mathrm{s})$$

$$= \sum_{\mathrm{r,n,s}}\ln \lambda \cdot \zeta_1(\mathrm{r}, \mathrm{n}, \mathrm{s}) = \sum_{\mathrm{r}}\ln \mathrm{r} \cdot \sum_{\mathrm{n}}\gamma_1(\mathrm{r}, \mathrm{n})$$

Consider $\mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|\mathrm{X}]$;

$$\mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|\mathrm{X}]$$

$$= \sum_{\mathrm{r,n,s}} \mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}, \mathrm{X}] \cdot P(\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}|\mathrm{X})$$

$$= \sum_{\mathrm{r,n,s}} \mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}] \cdot \zeta_1(\mathrm{r,n,s})$$

where $\mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|\ln\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}]$

$$= Var[\ln\lambda_1|\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}] + (\mathbb{E}[\ln\lambda_1|\lambda_1 = \mathrm{r}, \Delta N_1 = \mathrm{n}, \lambda_2 = \mathrm{s}] - \hat{a}_0)^2$$

$$= 0 + (\ln\mathrm{r} - \hat{a}_0)(\ln\mathrm{r} - \hat{a}_0)'$$

$$= (\ln\mathrm{r} - \hat{a}_0)(\ln\mathrm{r} - \hat{a}_0)'$$

$$\therefore \mathbb{E}[(\ln\lambda_1 - \hat{a}_0)(\ln\lambda_1 - \hat{a}_0)'|X] = \sum_{\mathrm{r,n,s}} (\ln\mathrm{r} - \hat{a}_0)(\ln\mathrm{r} - \hat{a}_0)' \cdot \zeta_1(\mathrm{r,n,s})$$

$$= \sum_{\mathrm{r}} (\ln\mathrm{r} - \hat{a}_0)(\ln\lambda - \hat{a}_0)' \cdot \sum_{\mathrm{n}} \gamma_1(\mathrm{r,n})$$

Consider $\mathbb{E}[u_t|\mathrm{X}]$;

$$\mathbb{E}[u_t|X] = \sum_{\mathrm{r,n,s}} \mathbb{E}[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}, \mathrm{X}] \cdot P(\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}|\mathrm{X})$$

$$= \sum_{\mathrm{r,n,s}} \mathbb{E}[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}] \cdot \zeta_{t-1}(\mathrm{r,n,s})$$

$$= \sum_{\mathrm{r,n,s}} \left(\hat{a} + \hat{B}\left(\hat{B} + \mathrm{n}\hat{\Omega}\right)^{-1}(\ln\mathrm{s} - \hat{a} - \mathrm{c}\cdot\ln\lambda - \mathrm{n}\hat{\kappa})\right) \cdot \zeta_{t-1}(\mathrm{r,n,s})$$

where the last equation comes from (A1).

Consider $\mathbb{E}[(u_t - \hat{a})(u_t - \hat{a})'|\mathrm{X}]$;

$$\mathbb{E}[(u_t - \hat{a})(u_t - \hat{a})'|\mathrm{X}]$$

$$= \sum_{\mathrm{r,n,s}} \mathbb{E}[(u_t - \hat{a})(u_t - \hat{a})'|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}, \mathrm{X}] \cdot P(\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}|\mathrm{X})$$

$$= \sum_{\mathrm{r,n,s}} \mathbb{E}[(u_t - \hat{a})(u_t - \hat{a})'|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}, \mathrm{X}] \cdot \zeta_{t-1}(\mathrm{r,n,s})$$

where $\mathbb{E}[(u_t - \hat{a})(u_t - \hat{a})'|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}, \mathrm{X}]$

$$= Var[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}] + (\mathbb{E}[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}] - \hat{a})(\mathbb{E}[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}] - \hat{a})'$$

$$= \hat{B} - \hat{B}\left(\hat{B} + \mathrm{n}\cdot\hat{\Omega}\right)^{-1}\hat{B} + (\mathbb{E}[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}] - \hat{a})(\mathbb{E}[u_t|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}] - \hat{a})'$$

Consider $\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})\,|\mathrm{X}\right]^{-1}$;

$$\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})\,|\mathrm{X}\right]$$

$$= \sum_{t=2}^{T} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})|\mathrm{X}\right]$$

$$= \sum_{t=2}^{T}\sum_{\mathrm{r,n,s}} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})|X, \lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_{\mathrm{t}} = s\right]$$
$$\cdot P(\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_{\mathrm{t}} = s|\mathrm{X})$$

$$= \sum_{t=2}^{T}\sum_{\mathrm{r,n,s}} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_{\mathrm{t}} = s\right]$$
$$\cdot \zeta_{t-1}(\mathrm{r,n,s})$$

$$= \sum_{t=2}^{T}\sum_{\mathrm{r,n}} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}diag(\ln\lambda_{t-1})|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_{\mathrm{t}} = s\right]$$
$$\cdot \gamma_{t-1}(\mathrm{r,n})$$

where $\mathbb{E}\left[\ln\lambda_{t-1}\,\hat{B}^{-1}\ln\lambda_{t-1}\,|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}\right] = diag(\ln\mathrm{r}) * \hat{B}^{-1} * diag(\ln\mathrm{r})$

Consider $\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a})\,|\mathrm{X}\right]$;

$$\mathbb{E}\left[\sum_{t=2}^{T} diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a})\,|\mathrm{X}\right]$$

$$= \sum_{t=2}^{T}\sum_{\mathrm{r,n,s}} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}(u_t - \hat{a})|\mathrm{X}, \lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}\right]$$
$$\cdot P(\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}|\mathrm{X})$$

$$= \sum_{t=2}^{T}\sum_{\mathrm{r,n,s}} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}(\mathrm{u_t} - \hat{\mathrm{a}})|\mathrm{X}, \lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}\right]$$
$$\cdot P(\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}|\mathrm{X})$$

$$= \sum_{t=2}^{T}\sum_{\mathrm{r,n,s}} \mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{B}^{-1}(\mathrm{u_t} - \hat{\mathrm{a}})|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}\right]$$
$$\cdot \zeta_{t-1}(\mathrm{r,n,s})$$

where $\mathbb{E}\left[diag(\ln\lambda_{t-1})\hat{\mathrm{B}}^{-1}(\mathrm{u_t} - \hat{\mathrm{a}})|\lambda_{t-1} = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_t = \mathrm{s}\right] = diag(\ln\mathrm{r}) *$
$\hat{B}^{-1} * \mathbb{E}[(u_t - \hat{a})|\lambda_t = \mathrm{r}, \Delta N_{t-1} = \mathrm{n}, \lambda_{t+1} = \mathrm{s}]\}$

Appendix 2: Other Graph Results

Figure 9: Return diffusion term for each countries data

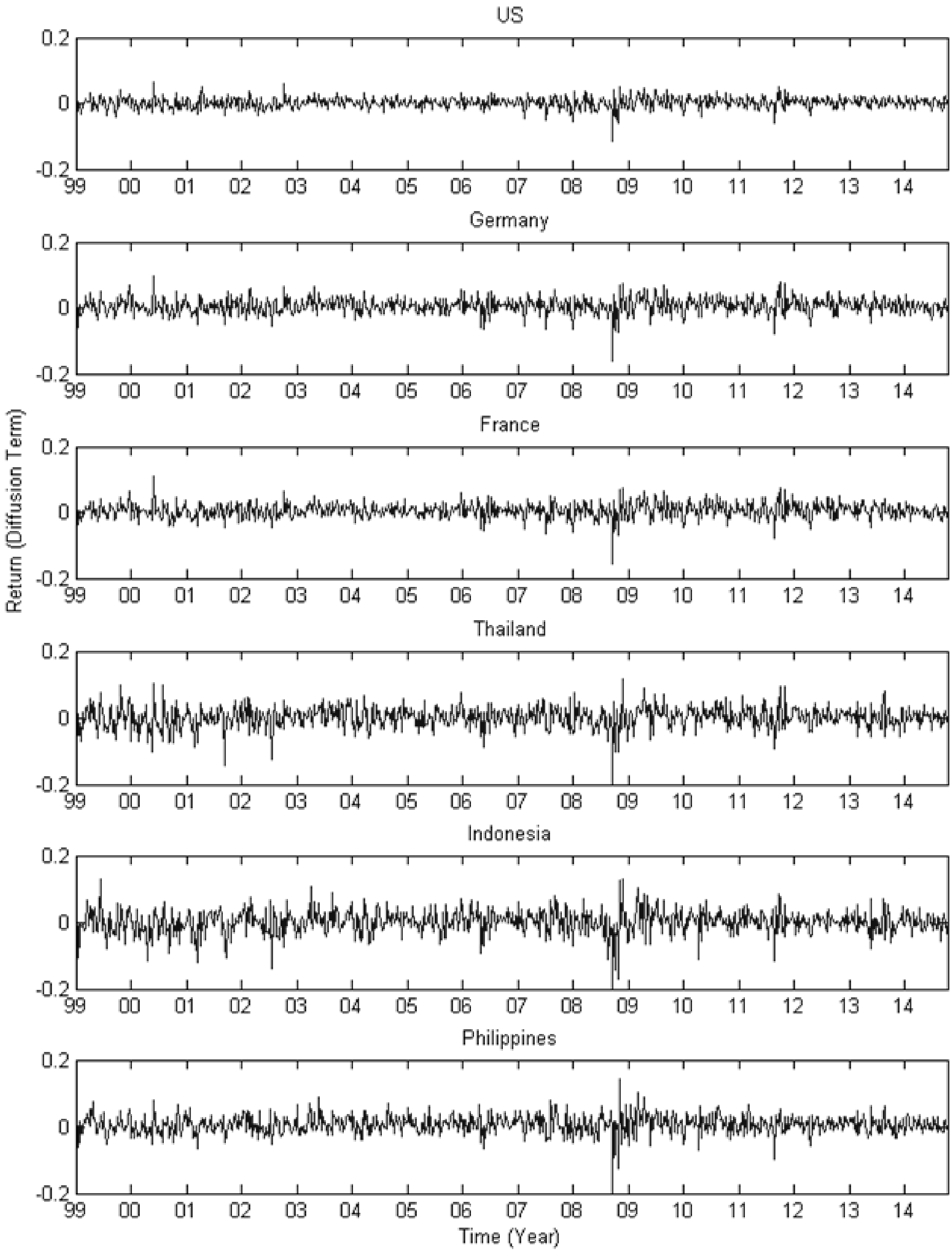
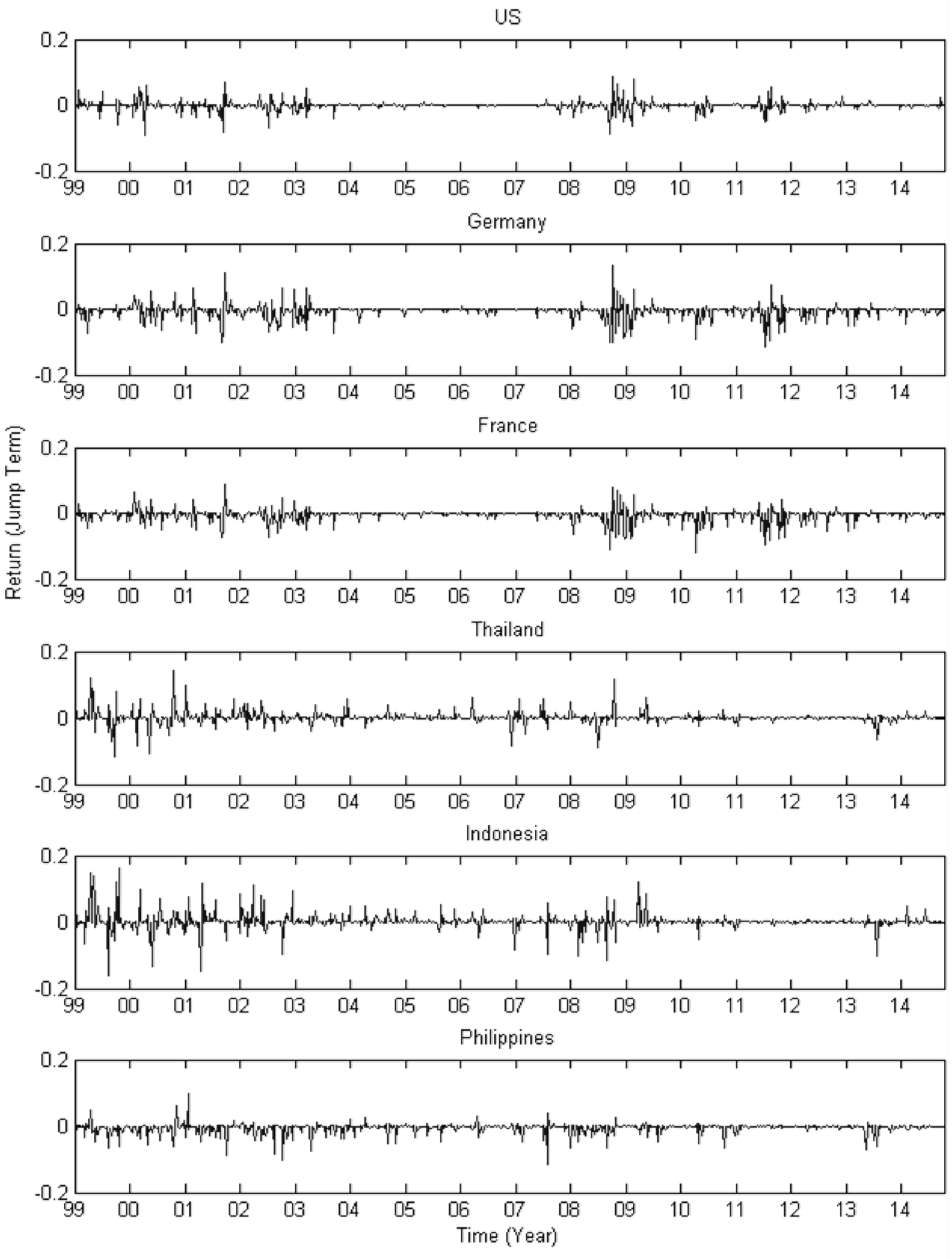


Figure 10: Return jump term for each countries data



Appendix 3: Information Matrix Approximation

$$I(\Gamma) = -\mathbb{E}\left[\frac{\partial^2}{\partial \Gamma^2} \log P(X;\Gamma) \,\middle|\, \Gamma\right]$$

Finite differences on derivative of bivariate functions according to Eberly (2003) is used to approximate the Hessian matrix $\frac{\partial^2}{\partial \Gamma^2} \log P(X;\Gamma)$.

Let $f(x, y)$ be the incomplete-data log-likelihood where x and y are the parameters in the Γ parameter's set.

Let $F(x+h, y+k)$ means the incomplete-data log-likelihood of when x plus some increment h and y plus some increment k while other parameters in the Γ's set remains the same.

The partial derivative centered differences approximation are

$$f_x(x, y) \approx \frac{F(x+h, y) - F(x-h, y)}{2h}$$

$$f_y(x, y) \approx \frac{F(x, y+k) - F(x, y-k)}{2k}$$

$$f_{xx}(x, y) \approx \frac{F(x+h, y) - 2f(x, y) + F(x-h, y)}{h^2}$$

$$f_{xy}(x, y) \approx \frac{F(x+h, y+k) - F(x+h, y-k) - F(x-h, y+k) + F(x-h, y-k)}{4hk}$$

So the hessian matrix of the incomplete-data log-likelihood would be

$$H(\Gamma) = \begin{bmatrix} f_{\Gamma_1\Gamma_1} & f_{\Gamma_1\Gamma_2} & \cdots & f_{\Gamma_1\Gamma_n} \\ & f_{\Gamma_2\Gamma_2} & & f_{\Gamma_2\Gamma_n} \\ & & \ddots & \vdots \\ & & & f_{\Gamma_n\Gamma_n} \end{bmatrix}$$

$\therefore I(\Gamma) = -H(\Gamma)$

The covariance matrix of the parameters is $I(\Gamma)^{-1}$

VITA

Yotsanan Simakorn was born in Bangkok, Thailand. He received a Bachelor of Engineering in Automotive Design and Manufacturing Engineering from International School of Engineering, Chulalongkorn University in June 2012. In August,2012, he entered Chulalongkorn Business School at Chulalongkorn University.