Bayesian Inference for Markov Modulated Levy Process

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Tridash Srivastava

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2024

Supervisor: Prof.Dr.Ludger Overbeck

Certificate

This is to certify that this dissertation entitled Bayesian Inference for Markov Modulated Levy Process towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents research carried out by Tridash Srivastavaat Indian Institute of Science Education and Research under the supervision of Prof.Dr.Ludger Overbeck, Professor, Department of Mathematics , during the academic year 2023-2024.

Prof.Dr.Ludger Overbeck

Committee:
Prof.Dr.Ludger Overbeck
Prof.Dr.Anindya Goswami

This thesis is dedicated to my Parents and Mentors

Declaration

I hereby declare that the matter embodied in the report entitled Bayesian Inference for Markov Modulated Levy Process are the results of the work carried out by me at the Department of Department of Mathematics, Indian Institute of Science Education and Research, Pune, under the supervision of Prof.Dr.Ludger Overbeck and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Tridash Srivastava{Roll no.20191226}

Acknowledgments

I am extremely grateful to Prof. Ludger Overbeck for supervising the project. This thesis has been possible because of his unwavering support and excellent guidance thorought the project. He has invested a lot his precious time in this project and made sure to clear all my doubts . He also secured the essential funding for me to visit his lab in Giessen, Germany as well as helped me greatly throughout the immigration process. I would also like to thank Prof. Anindya Goswami for his valuable comments and guidance in the project. I would also like to thank the Mathematical Institute, Justus Liebig University Giessen, for providing the facilities to carry out this project without difficulty.

I am also grateful to Prof. Matthias Meiners for allowing me to present my work at Oberseminar, JLU Giessen. I am also thankful to my parents, who inspired and encouraged me to excel at each step. I would also like to thank Anuradha for supporting me emotionally. I would also like to thank David, Florian and Arefeh for helping me navigate a new country and great conversations. I would also like to thank Miss. Julia Volz of the International Office at JLU for helping me through the moving process. I am also thankful to my friends Ravish, Yadram, Anitha, Vishal and Rakshit for helping me through this journey.

Abstract

Levy Process are used in finance for Asset Modeling and Risk Management. Markov Modulated Levy Process(MMLP) are a more flexible class of Stochastic Processes which capture phase changes arising in economies by allowing jumps in drift and volatility, linked to hidden states of a Markov chain. Theses models have been used to model option prices, renewable energy markets as well as for risk quantification. While Bayesian inference methods exists for simpler regime-switching models, we aim to extend it to more complex MMLPs.

Our approach involves applying Bayesian estimation techniques to recover the hidden states and the parameters associated with each state of the Markov Chain. We propose Markov Chain Monte Carlo algorithms to perform Bayesian inference for MMLPs. This will allow for a more data-driven analysis of asset returns with regime shifts and jumps.

Contents

Chapter 1

Introduction

Most of the financial models used assumes the market conditions to be uniform but that is usually not the case as economies undergoes changes in their state (e.g expansion,normal,recession). Regime switching models address this by allowing the model parameters to also switch between small number of distinct states. These states are described by a special type of random process called a Markov chain, which jumps between states according to fixed probabilities.

During each state the model behaves like a standard Levy process. Which implies the asset price has defined randomness and trend dependent on the state. When the underlying Markov chain state changes the randomness and the trend also switches. When a model combines these two elements - a Markov chain for state switching and a Levy process for price movement within each state - it's called Markov-modulated Levy process (MMLP). This model is more realistic than standard models because it captures the idea that market conditions can change over time. Hidden Markov Models or Regime switching models have been used extensively in finance literature to account for different type of phases occurring in the economy. [Cheng-Der Fuh, 2012] uses regime switching Brownian motion for pricing options and [Anindya Goswami, 2019] price options in a regime switching jump-diffusion models. Also, [Ivanov, 2022] explores risk quantification for Variance-Gamma Process with regime switching and [Rasmus, 2016] explores derivative pricing for regime switching levy process particularly regime switching Normal Inverse Gaussian(NIG) process. [Shaw, 2019] uses similar model to study US Corporate Option-Adjusted Spreads. Regime switching Levy process also find great use for modelling wind energy output, in the paper [Veraart, 2016] models the impact of wind production on electricity

prices using a regime-switching Levy process. Since, regime switching levy process are widely used in various fields including derivative pricing, risk quantification and energy prices modelling. Therefore a robust and efficient algorithm to carry out inference for these models is necessary.

In this project, we address the problem of inference for regime switching models for asset returns using Markov Modulated Levy processes (MMLP).The regime switching extension is one of many generalizations of Levy Process models of asset price dynamics. Bayesian Inference for Markov modulated geometric Brownian motion (MMGBM) models have already been explored e.g. in [Srikanth K. Iyer, 2009] so we naturally intend to extend it to Levy Models. The paper [Das and Goswami, 2019] by Milan Kumar Das and Anindya Goswami develops a statistical methodology for evaluating binary regime-switching models, specifically extensions of the geometric Brownian motion (GBM) model using squeeze duration analysis, We plan to explore possibility of applying Bayesian estimation techniques for estimating the parameters associated with Markov Modulated Levy Processes. [Milan Kumar Das, 2023] also explores binary regime models with jump discontinuities, focusing on inference methods and practical implications in their study. We wish to come up with Bayesian inference procedure for Markov modulated Levy processes, that allows the drift and volatility of the asset(Stock) to jump among different values. The jumps are determined by a continuous-time Markov chain but in a way that each state of the chain is associated with one value for drift and one for volatility. An MCMC algorithm is to be developed to perform exact Bayesian inference for MMLPs.

The general Regime switching Levy Process

Let X_t be a continuous time Markov chain on finite space. A regime-switching model is a stochastic process (S_t) which is solution of the stochastic differential

equation given by

$$dS_t = \kappa(X_t)(\eta(X_t) - S_t)dt + \sigma(X_t)dY_t^{\zeta(X_t)} \tag{1.1}$$

where $\kappa(X_t), \eta(X_t), \sigma(X_t)$ are functions of the Markov chain X.

- κ denotes the mean-reverting rate;
- η denotes the long-run mean;
- σ denotes the volatility of X.

$Y^{\zeta(X)}$ is a stochastic process which is a NIG Levy process with its parameters $(\mu, \delta, \alpha, \beta)$also Markov modulated.

To carry out Bayesian inference for the Markov Modulated Levy process, we consider the path of the Markov Chain as an additional parameter. If we consider M possible states of the Markov Chain then our parameter space would be $\hat{\Theta} = \{(\hat{\kappa_i}, \hat{\theta}_i, \hat{\sigma}_i, \hat{\mu}_i, \hat{\delta}_i, \hat{\alpha}_i, \hat{\beta}_i,) \quad \text{for} \quad i = 1, .., M\}$ and we would also have the path space $(X_t)_{t\in[0,n]} \in D([0, n], 1, ..., M)$. If we have observations $(S_1, .., S_n) = S_{1:n}$ and a given prior distribution for each of the parameters, then we have to find the posterior distribution of $\Theta, X_{[0,n]}$ conditional on our observation that is we try to estimate the density $f(\Theta; X_{[0,n]}|S_{1:T})$.

The Thesis is structured as follows. The first chapter introduces the problem statement and the motivation for carrying out this project. The second chapter introduces important concepts and definitions which will be used throughout the thesis. The third chapter introduces Bayesian Infernce using Truncated Dirichlet prior method for regime switching process in a genralized setting. The fourth chapter provides methedology for Truncated Dirichlet prior method of bayesian infernce for markov modulated Ornstein-Uhlbeck process. The fifth chapter provides methedology for Truncated Dirichlet prior method of bayesian infernce for markov modulated geometric Levy(NIG) process. The sixth chapter introduces Particle Markov Chain Monte Carlo method for bayesian infernce in generalized setting as well as provide methodology about how to use it for our two models. The seventh chapter contains results of numerical experiments and the eighth chapter includes discussion for implications of our work and the scope for future research.

Chapter 2

Preliminaries

We start by introducing essential concepts that are used throughout thesis. We first introduce Levy processes. Then we give a brief introduction on Bayesian Inference. Then we introduce Dirichlet prior and stick breaking process which are very important for our method. And at the end we give methods to estimate transition matrices form sample paths of the Markov Chain.

2.1 Levy Process

Levy process are the most used class of stochastic process to model financial data, ecological data as well as in signals processing. We first define them mathematically and also give important characterisations associated with them. The refrence used in this section are [Cont and Tankov, 2004],[Schoutens, 2003] and[Barndorff-Nielsen, 1997].

Definition 1. *A cadlag stochastic process(Y_t) on a given probability space is a Levy process if satisfies the following three conditions:*

- *Stationary increments : The distribution of $Y_{t+h} - Y_t$ doesn't depends on t*
- *Independent increments: for any increasing sequence of times $t_0 ... t_n$, the random variables $Y_{t_0}, Y_{t_1} - Y_{t_0}, ..., Y_{t_n} - Y_{t_{n-1}}$ are independent.*
- *Stochastic Continuity:*$\forall \epsilon > 0 \ \lim_{h \to 0} \mathbb{P}(|Y_{t+h} - Y_t| \geq \epsilon) = 0$

One thing to note that is that the third condition doesn't imply that sample paths are continuous.

Another important property of a Levy process (Y_t) is that its distribution is infinitely divisible.
Also if F is an infinitely divisible distribution then there exists a Levy process (Y_t) such that the distribution of Y_1 is given by F.
The simplest Levy processes are the compound Poisson process whose sample paths are piece-wise constant functions.

Definition 2. *We define the Levy measure for a Levy process as follows:*
let (Y_t) be a levy process on $\mathbb{R}^d$. The measure μ on $\mathbb{R}^d$ is:

$$\mu(A) = E[\#\{t \in [0,1] : \Delta Y_t \neq 0, \Delta Y_t \in A\}] \quad A \in \mathcal{B}(\mathbb{R}^d) \tag{2.1}$$

The most important results for Levy processes are the Levy Ito decomposition and Levy-Khinchin representation which gives us two ways of characterizing Levy processes.

Theorem 1. ***Levy Ito decomposition*** *Let (Y_t) be a Levy process on $\mathbb{R}^d$ and μ is its Levy measure , Then*

- *The jump measure of Y, denoted by J_Y, is a Poisson random measure on $[0,\infty]\times R^d$ with intensity measure $\mu(dy)dt$.*
- *There exist a vector γ and a d-dimensional Brownian motion (B_t) with covariance matrix A such that:*

$$Y_t = \gamma_t + B_t + Y_t^l + \lim_{\epsilon \to 0} \hat{Y}_t^{\epsilon} \tag{2.2}$$

The above theorem explains that every Levy process can be characterised by three quantities that are also called the Levy triplet of the process and include a vector γ, a positive definite matrix A and a positive measure μ that uniquely determine its distribution.

Theorem 2. ***Lévy–Khintchine representation*** *Let (Y_t) be a Levy process on $\mathbb{R}^d$ with Levy triplet (A,μ,γ) and if we have $E(e^{izY_t}) = e^{t\psi(z)}$ then,*

$$\psi(z) = -\frac{1}{2} z.Az + i\gamma.z + \int_{\mathbb{R}^d} (e^{izy} - 1 - izy1_{|y|\leq 1})\mu(dy) \tag{2.3}$$

Definition 3 (Subordinator)**.** *A Subordinator is a stochastic process* $Y = (Y_t)_{t\geq 0}$ *that is a Levy process and has the following property*

$$P(Y_t \geq 0 \quad \textit{for every} \quad t > 0) = 1$$

Proposition 1. *Let* $Y = (Y_t)_{t\geq 0}$ *be a real valued Levy process then the following are equivalent:*

1. $P(Y_t \geq 0 \quad \textit{for every} \quad t > 0) = 1$
2. $P(Y_t \geq 0 \quad \textit{for some} \quad t > 0) = 1$
3. *Sample paths are almost surely non decreasing i.e* $t \geq s \Longrightarrow Y_t \geq Y_s$ *with probability 1*
4. *Let (A,*μ*,b) be the charaterisitc triplet of* Y_t *then*

$$A = 0, \quad \mu((-\infty, 0]) = 0, \quad \int_0^\infty (x \wedge 1)\mu(dy) < \infty \quad \textit{and} \quad b > 0$$

To construct new Levy processes, we use three basic types of transformations, under which the class of Levy processes is invariant: linear transformations, subordination (time changing a Levy process with another increasing Levy process) and exponential tilting of the L´evy measure. Linear transformation is the easiest one which specifies that any linear transformation of a Levy process also gives a levy process.

The second method is **Subordination** which is also the method through which NIG process is developed which can be explained as follows:

Consider $(S_t)_{t\geq 0}$ as a subordinator and $(W_t)_{t\geq 0}$ as an independent Brownian motion. When we combine Brownian motion with a drift parameter μ using the subordination process S, we create a fresh Levy process denoted as $X_t = \sigma W(S_t) + \mu S_t$. If we examine this process on a different time scale, specifically the stochastic time scale defined by S_t, it transforms into a Brownian motion. This new time scale has a financial implication as it represents business time, indicating the cumulative rate at which information is received.

Barndorff-Nielsen in his paper [Barndorff-Nielsen, 1997] gives us the construction of the NIG process which we are going to be using in our project. It is constructed by using an Inverse Gaussian Process as a subordinator .

Definition 4 (Inverse Gaussian Process)**.** *Inverse Gaussian Process is a Levy process*

with the marginal distribution given by Inverse Gaussian distribution. Formally

$$Z^{(IG)} = \{Z_t^{(IG)}, t \geq 0\}$$

with parameters a and b >0 and has independent and stationary increments such that

$$\begin{aligned} Z_t - Z_s \sim & f_{IG}(z; a(t-s), b) \\ & = \frac{a(t-s)}{\sqrt{2\pi}} e^{a(t-s)b} z^{-3/2} e^{\frac{-1}{2}((a(t-s))^2 z^{-1} + b^2 z)} \quad z > 0 \end{aligned}$$

Here we use the definition provided by Ole Barndorff-Nielsen in his paper [Barndorff-Nielsen, 1997]
.

Definition 5 (NIG Process)**.** *if Z_t is an Inverse Gaussian Process(see [Cont and Tankov, 2004]) with parameter δ and $\sqrt{\alpha^2 - \beta^2}$ then we can represent NIG process as:*

$$Y_t = W(Z_t) + \beta Z_t + \mu t \tag{2.4}$$

Where W_t is a Brownian motion independent of z_t with drift 0 and diffusion coefficient 1.

The marginal distribution of the NIG process is called the NIG distribution which has four parameters $(\mu, \delta, \alpha, \beta)$.

- Location parameter (μ): location of the distribution. $\mu \in \mathbb{R}$
- Scale parameter (δ): scale/variance of the distribution. $\delta \in \mathbb{R}_{>0}$
- Shape parameter (α): tail heaviness. $\alpha \in \mathbb{R}_{>0}$
- Rate parameter (β): skewness of the distribution. $\beta \in \mathbb{R}$ and $-\alpha < \beta < \alpha$

The probability density function of NIG distribution can be written as

$$f_{\text{NIG}}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha}{\pi} \exp\left(\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)\right) \frac{K_1(\alpha\delta\sqrt{1 + (x-\mu)^2/\delta^2})}{\sqrt{1 + (x-\mu)^2/\delta^2}} \tag{2.5}$$

Here $K_1(x) = \frac{1}{2}\int_0^\infty exp(-\frac{1}{2}x(t + t^{-1}))dt$ denotes the third Bessel kind function with index 1. Therefore we can also define NIG process as

$$Y^{(NIG)} = \{Y_t^{(NIG)}, t \geq 0\}$$

with parameters $\alpha, \beta, \mu, \delta$ and has independent and stationary increments such that

$$Y_t - Y_s \sim f_{\text{NIG}}(x; \alpha, \beta, (t-s)\delta, (t-s)\mu)$$
$$= \frac{\alpha}{\pi} \exp\left((t-s)\delta\sqrt{\alpha^2-\beta^2} + \beta(x-(t-s)\mu)\right) \frac{K_1(\alpha(t-s)\delta\sqrt{1+(x-(t-s)\mu)^2/((t-s)\delta)}}{\sqrt{1+(x-(t-s)\mu)^2/((t-s)\delta)^2}}$$

We also observe that Normal distribution is just a special case of NIG distribution which can rises as a special case by setting $\beta = 0$ and $\delta = \sigma^2\alpha$ and letting $\alpha \to \infty$.Therefore we can also say that the Brownian motion is a subclass of NIG Levy process.

2.2 Bayesian Inference

Bayesian inference is a statistical method for updating probabilities based on new evidence. The refrence used in this section are [Haugh, 2021]. It involves the use of Bayes' theorem, which can be expressed as:

Theorem 3 (Bayes Theorem)**.** *Let A and B random variables in a measurable probability space, then*

$$f_{A|B}(a|b) = \frac{f_{B|A}(b|a) \cdot f_A(a)}{f_B(b)} \tag{2.6}$$

where:

$f_{A|B}(a|b)$ *is the probability density of A given $B = b$,*
$f_{B|A}(b|a)$ *is the conditional probability density function (PDF) of B given $A = a$,*
$f_A(a)$ *is the probability density function (PDF) of A,*
$f_B(b)$ *is the marginal probability density function (PDF) of B.*

In this continuous case, the probabilities are represented by probability density functions, and integrals are used instead of summations.

2.2.1 Bayesian Infernce

In Bayesian setting we consider that our parameters for the model as well as the data to be observed are random variables. We assign the parameter some probability density based on our initial knowledge which is called the prior density. Once the data

is observed we use Bayes Theorem (2.6) to update our prior density. The updated densities is called posterior density.

Definition 6 (Prior density)**.** *Let Θ be some unknown parameter vector of interest. We assume Θ is random with some prior density, $f_\Theta(\theta)$. This captures our prior uncertainty and knowledge regarding Θ. So we have initially :*

$$\Theta \sim f_\Theta(\theta)$$

We also consider our data to be random variable X and we consider observed data to be a realisation of X. Therefor we can define

Definition 7 (Likelihood)**.** *The likelihood is the conditional probability of observing the collected data x, given that we have the parameter to be $\Theta = \theta$. Represented mathematically as:*

$$f_{X|\Theta}(x|\theta)$$

Now, we want to get the posterior density which is the updated density of Θ after observing $X = x$ formally defined as

Definition 8 (Posterior density)**.** *The Posterior density is the conditional probability of Θ given that we have the data X. Represented mathematically as:*

$$f_{\Theta|X}(\theta|x)$$

The joint density of θ and X is given by

$$f(\theta, x) = f(x|\theta)f(\theta) \tag{2.7}$$

Now we use Bayes's Theorem to compute the posterior distribution from the prior and likelihood as follows

$$f_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta) \cdot f_\Theta(\theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta) \cdot f_\Theta(\theta)}{\int_\Theta f_{X|\Theta}(x|\theta) \cdot f_\Theta(\theta) d\theta} \tag{2.8}$$

Therefore using Bayes theorem we get a probability density over Θ, sometimes a point estimate of Θ is required then we use the following:

Definition 9 (maximum a posterior (MAP) estimator)**.** *The Mode of the posterior density or the θ for which $f_{\Theta|X}(\theta|x)$ is maximum is called the maximum a posterior (MAP) estimator.*

Much of Bayesian analysis is concerned with "understanding" the posterior $f(\theta|X)$. We can observe

$$f_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta) \cdot f_{\Theta}(\theta) \tag{2.9}$$

Where $\propto$ symbol denotes that $f(\theta|x) = C \times f_{X|\Theta}(x|\theta) \cdot f_{\Theta}(\theta)$ where C is a constant which doesn't depends on θ but might depend on X. Sometimes we can recognize the form of the posterior by simply inspecting $f_{X|\theta}(X|\theta) \cdot f_{\theta}(\theta)$. But typically we cannot recognize the posterior and cannot compute the denominator in (2.8) either. In such cases approximate inference techniques such as Markov Chain Monte Carlo are required.

2.3 Markov Chain Monte Carlo Sampling

2.3.1 The Sampling Problem

Suppose we are given a probability density

$$f(z) = \frac{\hat{f}(z)}{Zp}$$

where $\hat{f}(z) \geq 0$ is easy to compute but Z_p which is the normalization constant is (too) hard to compute. This very important situation arises in several contexts:
In Bayesian models such as ours where $\hat{f}(\theta) = f_{X|\theta}(X|\theta) \cdot f_{\theta}(\theta)$ is easy to compute but $Zp = \int_{\theta} f_{X|\theta}(X|\theta) \cdot f_{\theta}(\theta) d\theta$ can be very difficult or impossible to compute.

2.3.2 Metropolis-Hastings Algorithm

Coming to our sampling problem, suppose we want to sample from a probability density $f(z) = \frac{\hat{f}(z)}{Zp}$. To do this we construct a (reversible) Markov chain as follows and this method is called the Metropolis-Hastings Algorithm. Here's a step-by-step description of the algorithm:

1. Initialize the chain with an arbitrary starting point x_0.
2. For each iteration $t = 1, 2, 3, \ldots, T$:
   (a) Propose a new sample x' from a proposal distribution $q(x'|x_t)$.
   (b) Compute the acceptance ratio:

$$\alpha = \min\left(1, \frac{\hat{f}(x') \cdot q(x_t|x')}{\hat{f}(x_t) \cdot q(x'|x_t)}\right)$$

   (c) Set $X_{t+1} = x'$ with probability α otherwise set $X_{t+1} = x$

In the algorithm, $f(z)$ is the target distribution, $q(x'|x_t)$ is the proposal distribution, and T is the total number of iterations. The acceptance ratio α determines whether the proposed sample is accepted or rejected.

The resulting Markov chain is reversible with stationary distribution $f(z) = \frac{\hat{f}(z)}{Zp}$. We can therefore sample from f(z) by running the algorithm until stationarity is achieved and then using generated points as our samples. Note that Zp is not required for the algorithm! Therefore to sample from the posterior of any parameter θ , $f(\theta|X)$ we need $f_{X|\theta}(X|\theta) \cdot f_\theta(\theta)$ only as given in 2.9 using Metropolis Hastings Algorithm.

2.3.3 Gibbs Sampling

We use in this subsection [Tierney, 1994] as a reference. Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm used for generating samples from a multivariate probability distribution. It is particularly useful when it is difficult to directly sample from the joint distribution of all variables but is relatively easy to sample from the conditional distributions of individual variables given the values of the other variables.

- We want to sample from random variable x=$(x_1, .., x_d) \in \mathbb{R}$ with distribution π(x$_1, .., x_d$)
- we can simulate the distribution of each component conditional on the others, i.e. we can draw from $\pi(x_k|x_1, .., x_{k-1}, x_{k+1}, .., x_d)$ for each k=1 to d.
- We want to sample from the joint distribution, π(x). Gibbs sampling constructs a path of Markov Chain, $x^{(1)} \to x^{(2)} \to x^{(3)}$ with each step given by
  Simulate

- $x_1^{j+1}$from$\pi(x_1^{j+1}|x_2^j, ..., x_d^j)$
- $x_2^{j+1}$from$\pi(x_2^{j+1}|x_1^{j+1}, x_3^j, ..., x_d^j)$
- $x_3^{j+1}$from$\pi(x_3^{j+1}|x_1^{j+1}, x_2^{j+1}, x_4^j, ..., x_d^j)$
- .....

Convergence of Gibbs sampling

Let X be a Markov chain with state space $\Omega \subseteq \mathbb{R}^d$ with the transition Kernel:

$$P(x, A) = Pr(X_{i+1} \in A|X_i = x) \quad x \in \Omega, \quad A \subset \Omega$$

Definition 10. *Markov Chain X is π^* irreducible if*

$$\pi^*(A) > 0 \Longrightarrow P(X_i \in A|X_0 = x) > 0 \ \ \textit{for every } x \ \in \Omega$$

for some $i \geq 1$.
This implies Markov Chain can visit all the states that have positive probability in π^* starting from any state in the state space.

Definition 11. *We say π is an invariant distribution for the above defined Markov Chain if :*

$$\pi P = \pi$$

We use the following proposition to show our Gibbs Sampler converges to the joint distribution

Proposition 2. *If P(.,.) is a π^* irreducible and has invariant distribution π^*, then π^* is the unique invariant distribution of P. If P(.,.) is also aperiodic then for almost every $x \in \Omega$ and all sets A we have*

$$|P^m(x, A) - \pi^*(A)| \to 0$$

as $m \to \infty$

Therefore to show that our Gibbs sampler Markov Chain converges to the true distribution we just have to show that

Proposition 3. *$\pi(x_1, .., x_d) = \pi(x)$ is the invariant distribution measure for Markov Chain created by gibbs sampler.*

Proof. The Markov chain created by the Gibbs sampler is a d-dimensional Markov Chain where each state change occurs as follows:

$$(x_1^i, ..., (x_d^i) \to (x_1^{i+1}, ..., (x_d^{i+1})$$

Therefore we can give the transition kernel as follows:

$$\begin{aligned} P(x, dy) &= \pi(y_1|x_2, ..., x_d) \times \pi(y_2|y_1, x_3, ..., x_d) \times ... \times \pi(y_d|y_1, ..., y_{d-1}) dy_1 ... dy_d \\ &= \prod_k \pi(y_k|y_1, ..., y_{k-1}, x_{k+1}, ..., x_d) \end{aligned}$$

Now, we first prove the following statement

$$\pi(y_k|y_1, ..., x_{k+1}, ..., x_d) = \frac{\pi(y_k|y_1, ..., y_{k-1})\pi(x_{k+1}, ..., x_d|y_1, ..., y_k)}{\pi(x_k + 1, ..., x_d|y_1, ..., y_{k-1})}$$

To the prove the above statement we use definition of conditional distribution as follows:

$$\begin{aligned} RHS &= \frac{\pi(y_k|y_1, ..., y_{k-1})\pi(x_{k+1}, ..., x_d|y_1, ..., y_k)}{\pi(x_k + 1, ..., x_d|y_1, ..., y_{k-1})} \\ &= \{\frac{\pi(y_1, ..., y_{k-1}, y_k)\pi(x_{k+1}, ..., x_d, y_1, ..., y_k)}{\pi(y_1, ..., y_{k-1})\pi(y_1, ..., y_k)}\}\{\frac{\pi(y_1, ..., y_{k-1})}{\pi(x_{k+1}, ..., x_d, y_1, ..., y_{k-1})}\} \\ &= \frac{\pi(x_{k+1}, ..., x_d, y_1, ..., y_k)}{\pi(x_{k+1}, ..., x_d, y_1, ..., y_{k-1})} \\ &= \pi(y_k|y_1, ..., y_{k-1}, x_{k+1}, ..., x_d) = LHS \end{aligned}$$

Now we finally prove our proposition that $\int P(x, dy)\pi(x)dx = \pi(y)$.

$$\begin{aligned} &\int P(x, dy)\pi(x)dx \\ &= \int ... \int \prod_k \pi(y_k|y_1, ..., y_{k-1}, x_{k+1}, ..., x_d)\pi(x_1, ..., x_d)dx_1 ... dx_d \quad \text{use above statement} \\ &= \int ... \int \prod_k \frac{\pi(y_k|y_1, ..., y_{k-1})\pi(x_{k+1}, ..., x_d|y_1, ..., y_k)}{\pi(x_k + 1, ..., x_d|y_1, ..., y_{k-1})} \pi(x_1|x_2, .., x_d)\pi(x_2, .., x_d)dx_1 .. dx_d \\ &= \prod_k \pi(y_k|y_1, ..., y_{k-1}) \int .. \int \prod_k \frac{\pi(x_{k+1}, ..., x_d|y_1, ..., y_k)}{\pi(x_k + 1, ..., x_d|y_1, ..., y_{k-1})} \pi(x_1|x_2, .., x_d)\pi(x_2, .., x_d)dx_1 .. dx_d \end{aligned}$$

Now observe $\prod_k \pi(y_k|y_1, ..., y_{k-1}) = \pi(y)$, So we just need to prove rest of terms equal

to one. Rest of the terms are

$$\begin{aligned}
&\int .. \int \frac{\pi(x_2, .., x_d|y_1)..\pi(x_d|y_1, .., y_{d-1}}{\cancel{\pi(x_2, .., x_d)}..\pi(x_d|y_1, .., y_{d-2})} \pi(x_1|x_2, .., x_d) \cancel{\pi(x_2, .., x_d)} dx_1..dx_d \\
&= \int .. \int \prod_j \frac{\pi(x_{j+1}, .., x_d|y_1, .., y_j)}{\pi(x_{j+2}, .., x_d|y_1, .., y_j)} \pi(x_1, .., x_d) dx_1, .., dx_d \\
&= \int .. \int \prod_j \pi(x_j|x_j + 1, .., x_d, y_1, .., y_j) dx_j \\
&= \int \pi(x|y) dx = 1
\end{aligned}$$

Hence we have proved that the joint distribution $\pi(x_1, .., x_d)$ is the invariant distribution for the markov chain created by the gibbs sampler

□

In our case we are going to use Gibbs Sampler to sample the joint distribution of the parameters using the full conditionals of the each variable of the parameter space.

2.4 Dirichlet process

Dirichlet process is a stochastic process whose sample paths are probability distributions. As we know stochastic processes are distributions over functions and paths of the stochastic process are random functions. For DP, it is a probability measures which are random functions with some properties and can be interpreted as probability distributions over some set S. Hence a single draw from a DP outputs distribution rather than producing a single parameter (vector).It is important to note that sample paths of DP are discrete distributions over set S with probability 1. A base distribution H and the concentration parameter, also known as the scaling parameter and denoted by the positive real number α, together define the Dirichlet process. As described in [Teh, 2010] The Dirichlet can be formally defined as follows:

Definition 12 (Dirichlet Process)**.** *A random distribution G is distributed according to a DP, if its marginal distribution is Dirichlet distributed. let H be a distribution over S and α be a positive real number. Then for any finite measurable partition $A_1, ..., A_r$ of S ,we say G is Dirichlet process distributed with base distribution H and*

concentration parameter α*, written* $G \sim DP(\alpha, H)$*, if*

$$(G(A_1), ..., G(A_r)) \sim Dir(\alpha H(A_1), ..., \alpha H(A_r))$$

for every finite measurable partition $A_1, ..., A_r$ *of* S*.*
Where "Dir" denotes the Dirichlet distribution.

Proposition 4. *Draws from Dirichlet Process are random measures over the set* S *that are discrete with probability one and therefore can be represented in the form:*

$$\mu(\omega, s) = \Sigma_{k=1}^{\infty} P_k(\omega)\delta_{s_k(\omega)}(s)$$

Where $s_k \in S$ *and* P_k *are random weights .*

The base distribution represents the expected value of the process; hence, the Dirichlet process draws distributions "around" the base distribution in a similar manner to how a normal distribution draws actual numbers around its mean.The scaling parameter determines the degree of this discretization. Since α defines the discretization we will have that as $\alpha \longrightarrow 0$, all realisations will be concentrated at a single value. Whereas when $\alpha \longrightarrow \infty$, we will have $G(A) \longrightarrow H(A)$ for any measurable A, that is $G(A) \longrightarrow H(A)$ pointwise. However this not equivalent to saying that $G \longrightarrow H$. As draws from a DP will be discrete distributions with probability one, even if H is smooth.

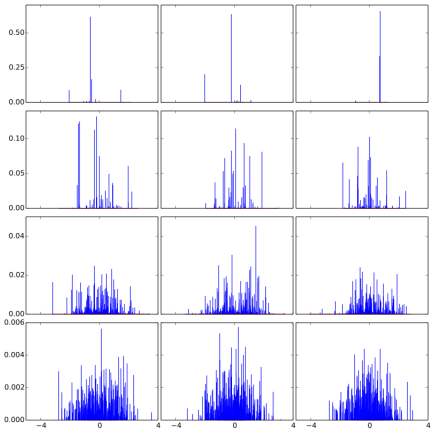


Fig. 2.1: Draws from the Dirichlet process $DP(N(0, 1), \alpha)$ **. The four rows use different** α **(top to bottom: 1, 10, 100 and 1000) and each row contains three repetitions of the same experiment**

2.4.1 Stick Breaking Process

Stick Breaking Process is one of the most common way to sample from Dirichlet process. It uses the important property of DP that sample paths of DP are discrete distributions over set S with probability 1. [Sethuraman., 1994] used this property to give a precise and straightforward construction of DP as follows:

Definition 13. *let $s_k \in S$, $k = 0$ to ∞ and $s_k(\omega)$ are independent and identical drawn according to H.*
And let $P'_k(\omega)$ be independently drawn from the distribution Beta (1,α) and let

$$P_k(\omega) = P'_k(\omega)\Pi_{i=1}^{k-1}(1 - P'_i(\omega))$$

Then we can define a random measure on S which is a realization of DP ,$\mu(\omega, s)$ as follows:

$$\mu(\omega, s) = \Sigma_{k=1}^{\infty} P_k(\omega)\delta_{s_k(\omega)}(s)$$

The resemblance to 'stick-breaking' can be seen by considering that we start with a unit-length stick and then break it at P'_1 and assigning P_1 the length of stick we broke off and in the next each step we break off a portion of the remaining stick according to P'_1 and assign this broken-off piece to P_1 . The idea behind this is to repeatedly cut off and throw away a random fraction (selected from a Beta distribution) of a "stick" that is initially 1 length. We explicitly employ the discreteness and provide the probability mass function of this (random) discrete distribution.

2.4.2 Finite truncation

We use an almost sure truncation of DP(α,H) in our algorithm like the one used in [Ishwaran and Zarepour, 2000] which can be defined as:

Definition 14 (Truncated Dirichlet process)**.** *let $s_k \in S$, $k = 0$ to N and $s_k(\omega)$ are independent and identically distributed according to H.*
Then the druncated dirichlet process will be of the following form

$$f_N(\omega, .) = \sum_{k=1}^{N} P_k(\omega)\delta_{s_k(\omega)}(.)$$

where,

$$
\begin{aligned}
&P_1(\omega) = P'_1(\omega) \\
&P_k(\omega) = P'_k(\omega)\Pi_{i=1}^{k-1}(1 - P'_i(\omega)) \quad (k = 2, ..., N-1) \\
&P_N(\omega) = \Pi_{i=1}^{N-1}(1 - P'_i(\omega))
\end{aligned}
$$

And $P'_k(\omega)$ *,* $k = 1, ..., N-1$ *are independent draws from the distribution Beta (1,*α*).*

2.4.3 Distribution of Proportions

In the following section we derive the probability distribution of P=$[P_1, ..., P_N]$ which will later allow us to get full conditional of P. The reference used in the section is [Connor and Mosimann, 1969].

Definition 15. *Let* $P_1, P_2, ..., P_k$ *be non negative continuous random variable satisfying the constraint* $\sum P_i = 1$*. Then P's are called* ***proportions****.*
We also define the following notations:

1. $S_j = \sum_{i=1}^{j} P_i$
2. $P'_i = P_i/[1 - S_{i-1}]$ *where* $P'_1 = P_1$ *and* $P'_k = 1$
3. $\boldsymbol{P}$= $(P_1, ..., P_k)$ *with* $P_{j1} = (P_1, ..., P_j)$ *and* $P_{j2} = (P_{j+1}, ..., P_k)$
4. $W_j = (1/[1 - S_j])P_{j2}$

Lemma 1. *Using the above definition we can see that* $P_i = P'_i[\prod_{m=1}^{i-1}(1 - P'_m)]$

Definition 16. *Given a random vector of proportions the proportion* P_1 *is said to be neutral if* P_1 *is independent of the vector* $(P'_i = P_i/[1 - S_{i-1}]; i \geq 2)$

Definition 17. *Given* $\boldsymbol{P}$ *divided such that* $\boldsymbol{P}$= (P_{j1}, P_{j2}) *.*P_{j1} *is a neutral vector if it is independent of* W_j*.*
If P_{j1} *is a neutral vector for all j, then* $\boldsymbol{P}$ *is said to be completely neutral.*

Theorem 4. *Suppose P is completely neutral . Then the random variables* P'_i *are mutually independent. Let density of each* P'_i *be uni-variate beta distribution* $B(a_i, b_i)$

then the density of P's are:

$$[\prod_{i=1}^{k-1}\mathcal{B}(a_i,b_i)]p_k^{b_{k-1}-1}\prod_{i=1}^{k-1}[p_i^{a_i-1}(\sum_{j=i}^{k}p_j)^{b_{i-1}-(a_i+b_i)}] \tag{2.10}$$

We call that the distribution as a generalized dirichlet distribution and represent it as $\mathcal{G}(a_1,b_1,...,a_k,b_k)$

Proof. We first see that the distribution of vector

$$\mathbf{P'} = (P_1',...,P_{k-1}') = [\prod_{i=1}^{k-1}\mathcal{B}(a_i,b_i)^{-1}z_i^{a_i-1}(1-z_i)^{b_i-1}]$$

Now we know that transformation of probability distribution of a random vector(X) to another random(Y) vector which is a function of the first one can be done by using the below formula.

$$g(Y) = |J|f(v(Y))$$

where v is the inverse of the function that maps X to Y and is the jacobian matrix for v .

Therefor in our case we have the jacobian to be $\begin{bmatrix} 1 & 0 & 0... \\ 0 & 1/[1\text{-}S_1] & 0... \\ 0 & 0 & 1/[1\text{-}S_2]... \\ . & . & . \end{bmatrix}$ therefore the Jacobian comes out to be $\prod_{i=1}^{k-1}[1/(1-S_{i-1})]$. Now applying the transformation we obtain the density function of $P = (p_1,..,p_{k-1})$ to be

$$[\prod_{i=1}^{k-1}\mathcal{B}(a_i,b_i)]p_k^{b_{k-1}-1}\prod_{i=1}^{k-1}[p_i^{a_i-1}(\sum_{j=i}^{k}p_j)^{b_{i-1}-(a_i+b_i)}]$$

□

2.5 Estimation for Markov Chains

In this section we introduce the methhods to estimate Markov Chain transition matrices if we have sample paths of the MC. This method will be later used as part of our method. The references used in this section are [Inamura, 2006] and [Shalizi, 2009].

2.5.1 Discrete time Markov chain with Discrete observations

Let X be a Discrete time Markov Chain with finite state space S=$(s_1, s_2, ..., s_K)$. We know that the a DTMC is defined by its transition matrix

$$\mathbb{P} = \begin{bmatrix} P_{11} & P_{12} & ... & P_{1K} \\ P_{21} & P_{22} & ... & P_{2K} \\ P_{31} & P_{32} & ... & P_{3K} \\ . & . & . & . \\ P_{K1} & P_{K2} & ... & P_{KK} \end{bmatrix}$$

where $P_{ij} = \text{Prob}(X(t_{n+1}) = j | X(t_n) = i)$ Here we describe the Maximum Likelihood method to get transition matrix from sample paths for a given DTMC.
Suppose we observe a sample path of the markov chain given as $x_1, x_2, ..., x_n$.
The probability of this realization is:

$$\text{Prob}(X_1 = x_1) \prod_{t=2}^{n} \text{Prob}(X_t = x_t | X_{t-1} = x_{t-1})$$

Which can be rewritten in terms of the transition probabilities P_{ij},

$$\begin{aligned} L(P) =& \text{Prob}(X_1 = x_1) \prod_{t=2}^{n} P_{x_{t-1} x_t} \\ =& \text{Prob}(X_1 = x_1) \prod_{i=1}^{k} \prod_{j=1}^{k} P_{ij}^{n_{i,j}} \end{aligned}$$

Then we take the log likelihood of the above expression:

$$log(L)(P) = \log(\text{Prob}(X_1 = x_1)) + \sum_{i,j} n_{i,j} \log(P_{i,j})$$

Now we try to find the maximum likelihood but we also need to be careful of the constraint

$$\sum_{j} P_{i,j} = 1$$

Since we have K constraint equations we use introduce K Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_K$. And we get the new objective function to be :

$$log(L)(P) - \sum_{i=1}^{j} \lambda_i (\sum_{j} P_{i,j} - 1) \tag{2.11}$$

Now taking derivative with respect to $P_{i,j}$ gives:

$$\begin{aligned} 0 =& \frac{n_{ij}}{P_{ij}} - \lambda_i \\ \lambda_i =& \frac{n_{ij}}{P_{ij}} \\ P_{ij} =& \frac{n_{ij}}{\lambda_i} \end{aligned}$$

from the constraint equation we have

$$\sum_{j=1}^{m} \frac{n_{ij}}{\lambda_i} = 1$$

$$\sum_{j=1}^{m} n_{ij} = \lambda_i$$

Therefore we have the **Maximum Likelihood Estimator** to be:

$$P_{ij}^{MLE} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}} \tag{2.12}$$

2.5.2 Continuous time Markov Chain with Continuous observations

We now move to a CTMC with continuous observations.Suppose we have A CTMC $X(t)$ with K states then it is characterized by its generator matrix or also called q

matrix.

$$\mathbb{Q} = \begin{bmatrix} Q_{11} & Q_{12} & ... & Q_{1K} \\ Q_{21} & Q_{22} & ... & Q_{2K} \\ Q_{31} & Q_{32} & ... & Q_{3K} \\ . & . & . & . \\ Q_{K1} & Q_{K2} & ... & Q_{KK} \end{bmatrix}$$

which has the following properties:

$$\sum_{j=1}^{K} Q_{ij} = 0 for 1 \leq i \leq K$$

$$0 \leq -Q_{ii} = \sum_{j \neq i} Q_{ij}$$

$$Q_{i,j} \geq 0 \text{for all } i \neq j$$

Now suppose we have sample path of the CTMC as X_t which includes all the instance in time when change in state occurs as well as the corresponding state change. Then considering that the CTMC changes state from i to j at t_1 and then from j to k at time t_2 and so on. We can then write the likelihood in the following way:

$$\begin{aligned} \mathbb{L}(Q) &= \exp(-Q_i(t_2 - t_1))Q_{ij} \exp(-Q_j(t_3 - t_2))Q_{jk}... \\ &= \prod_{i=1}^{K} \prod_{i \neq j} (Q_{ij})^{N_{ij}(T)} \exp(-Q_i R_i(T)) \end{aligned}$$

where $R_i(T) = \int_0^1 1_{\{x(s)=i\}} ds$ is the holding time at state i by the time t. $N_{ij}(t)$ is the number of times state changes from i to j by time t. We now take log of the above expression to get log likelihood as

$$\log(L)(Q) = \sum_{i=1}^{K} \sum_{j \neq i} \log(Q_{ij}) N_{ij}(T) - \sum_{i=1}^{K} \sum_{j \neq i} Q_{ij} R_i(T)$$

Hence the maximum likelihood estimator is:

$$Q_{ij}^{MLE} = \frac{N_{ij}(T)}{R_i(T)} \tag{2.13}$$

Chapter 3

Bayesian Inference for Regime Switching Processes

Now we will introduce Markov Modulation or regime switching process: We will first define a Markov-Switching as:

Definition 18. *Let $(X_t)_{t\in[0,T]}$ be a continuous time Markov chain on finite space S: = 1, 2, ..., M. The generator matrix of M, denoted by Π^M, is given by*

$$\Pi^M_{ij} \geq 0 \text{ if } i \neq j \text{ for all } i,j \in S \text{ and } \Pi^M_{ii} = -\sum_{j\neq i} \Pi^M_{ij} \text{ otherwise} \tag{3.1}$$

Next we consider some observable process say S which has independent and stationary increments and has parameters θ which determine its distribution. Mathematically,

$$S_{t+\tau} - S_t \sim f(\theta, \tau) \tag{3.2}$$

Where $f(\theta, \tau)$ denotes the density of the increments. Now we want the parameters denoted by θ to be regime switching or more precisely each state of the above defined continuous time Markov chain corresponds to a different set of parameters θ_k. If lets say the dimension of θ is K, i.e. the distribution f is a function of K variables, and

our Markov Chain has say M states. Then we can denote it as

$$\Theta : \begin{bmatrix} 1 \\ \vdots \\ M \end{bmatrix} \longrightarrow \left\{ \begin{bmatrix} \theta_{11} \\ \vdots \\ \theta_{1K} \end{bmatrix}, ..., \begin{bmatrix} \theta_{M1} \\ \vdots \\ \theta_{MK} \end{bmatrix} \right\} \tag{3.3}$$

where

$$\Theta(X_t) \longrightarrow \begin{bmatrix} \theta_{X_t 1} \\ \vdots \\ \theta_{X_t K} \end{bmatrix} \tag{3.4}$$

So we can observe that θ_{ij} corresponds to value of the j^{th}parameter of the density in state i.
So now our model can be denoted in the mathematical form as:

$$S_{t+\tau} - S_t \sim f(\Theta(X_t), \tau)$$
$$\text{If the Markov chain remains in state X(t) till } t + \tau$$

Note that now the distribution of increments is not identical or stationary anymore but still independent.
Now, we want to fit a regime-switching model such as above. Thus the optimal set of parameters to estimate is $(\hat{\theta}_{i,j}\, \hat{X(.)}) \in \mathbb{R}^{M\times K} \times \mathcal{D}$. where $i = 1, ..., M$ and $j = 1, ..., K$. Where $\{\hat{\theta}_{i,j}; i = 1, ..., M; j = 1, ..., K\} \in \mathbb{R}^{M\times K}$ are the parameters for the density associated and $\hat{X(.)} \in \mathcal{D}([0, T], \{1, ..., M\})$ is the path taken by the underlying Markov Chain and $\mathcal{D}$ is the set of Cadlag paths of the continuous time Markov Chain
We define a prior density p_θ for θ and then Bayesian inference relies on the joint density

$$p(\Theta; X_{[1,T]}|S_{[1,T]}) \propto p_\theta(X_{[1,T]}|S_{[1,T]})p(\theta)$$

For non-linear non-Gaussian models, $p_\theta(X_{[1,T]}|S_{[1,T]})$ and $p(\theta; X_{[1,T]}|S_{[1,T]})$ do not usually admit closed form expressions, making inference difficult in practice. It is therefore, necessary to resort to approximations. Monte Carlo methods have been shown to provide a flexible framework to carry out inference in such models.

3.1 Truncated Dirichlet Prior Method

Now, to carry out inference for our model we develop a new type of MCMC method called "Truncated Dirichlet Prior Method" which is inspired by similar methods in [Srikanth K. Iyer, 2009] and [Ishwaran and Zarepour, 2000] but this formulation for generalized regime switching process is given in this thesis.
We first think of each individual path taken by then Markov Chain as a class, to which we have to assign our observed data. Our models can be described in the following hierarchical form:

$$\begin{aligned}(S|X_t^i,\theta) \sim & f(S|X_t^i,\theta)\\ (X_t^i|P) \sim & P\\ \theta \sim & f(\theta)\\ P \sim & \mathcal{P}\end{aligned}$$

In this semi parametric setting, $S = (S_1, ..., S_n)$ is the observed data while $X = (X^1, ..., X^N)$ are the individual path taken by then Markov Chain and each $X_t^i \in \mathcal{D}([0,T]\{1, ..., M\})$ where $\mathcal{D}$ is the set of Cadlag paths of the continuous time Markov Chain or the path space of the Markov Chain. Here $\theta = (\theta_1, ..., \theta_M)$ is as defined above and each $\theta_i \in \mathbb{R}^K$

Now to carry out inference we place a Truncated Dirichlet process prior D(γ ,H), on the path space of the MC (X_t), with precision parameter γ and mean H which is a probability measure governing a MC on the path space D([0, ∞), S), the set of cadlag functions. The initial distribution according to H is the uniform distribution π_0 = (1/M, . . . , 1/M), and the transition matrix is Q with q_{ij} = 1/(M -1), i $\neq$j.

We can now rewrite the model as follows:

$$\begin{aligned}(S|X_i,\theta) \sim & f(S|X_i,\theta)\\ (X_i|p) \sim & \sum_{i=1}^{N} p_i \delta_{X_i}\\ (p,X) \sim & (p) \times H^N(X)\\ \theta \sim & (\theta)\end{aligned}$$

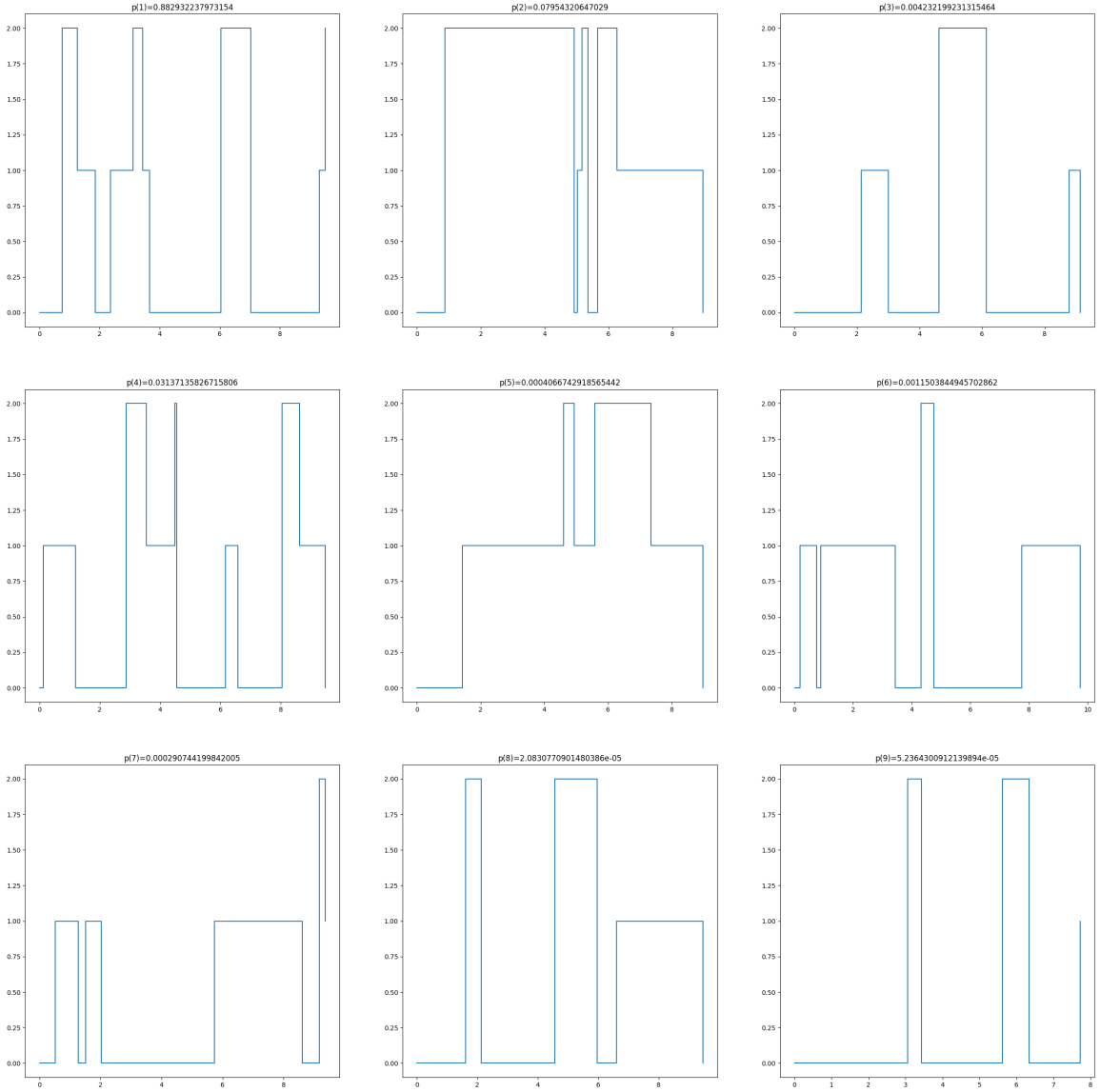


Fig. 3.1: Probability distribution over path space where we have chosen N=9 and M=3

where $X = (X^1, ..., X^N)$ are random paths generated from the distribution H and $(p_1, ..., p_N) \sim \mathcal{D}(1, \gamma)$ where p_i is generated from the stick breaking mechanism.
The algorithm works in the following ways:

- We generate a large number (N) number of paths of the discrete Markov chain with finite(M) number of states and treat them as classes.
- We use stick breaking mechanism to place a random probability distribution over our path space.
- We also place appropriate prior distribution over all the parameters modulated by the Markov Chain.
- Then we use the Blocked Gibbs Sampling technique to infer both the posterior distribution over the path space of the Markov Chain as well as the posterior distribution of the parameters simultaneously.

Blocked Gibbs Sampler
Once we have placed the appropriate priors for θ and X ,We use the Gibbs sampler to sample the posterior distribution $\mathcal{P}(\theta, X, p|\mathbf{S})$ directly. This method works by iteratively drawing values from the conditional distributions of the blocked variables

$$f(\theta|X, S)$$
$$f(X|\theta, p, S)$$
$$f(p|X)$$

Doing so eventually produces values drawn from the distribution of $(X, \theta, p|S)$.
To run the blocked Gibbs, draw values in the following order:

1. Conditional for each component of θ i.e. θ_{ij}:
   Draw value from the conditional

$$(\theta_{ij}|S, X) \sim f(\theta_{ij}) \prod_{k:X_{t_{k-1}}=i} \{f(S_k|\theta_{ij}, \theta_{i1}, ..., \theta_{i(j-1)}, \theta_{i(j+1)}, ..., \theta_{iK}, X, p)\}$$

2. Conditional for X :Draw values

$$f(X = x|p, \theta, S) \sim \sum_{i=1}^{N} p_i^* \delta_{X_i}(.)$$

$$p_i^* \propto \prod_{j=1}^{m} \{ \prod_{k:X^i_{t^i_{k-1}}=j} f(S_k|\theta_{j1}, ..., \theta_{jK}) \} p_i$$

3. Conditional for p
   By the conjugacy of the generalized Dirichlet distribution to multinomial sampling, it follows that our draw is
   $P = (p_1, ..., p_N)$

$$p_1 = V_1^* \quad and \quad p_k = (1 - V_1^*)...(1 - V_{k-1}^*)V_k^* \tag{3.5}$$

   where

$$(V_k^*|X_j) \sim \beta(a_k^*, b_k^*) \tag{3.6}$$

$$a_k^* = \begin{cases} a_k^* = a_k + 1 & if j = k \\ a_k^* = a_k & \text{else if } j \neq k \end{cases}$$

$$b_k^* = \begin{cases} b_k^* = b_k + 1 & if j > k > 0 \\ b_k^* = b_k & else if j \leq k \end{cases}$$

The resulting Algorithm is as follows:

1. Choose appropriate priors for each component of the parameter θ.
2. Generate p=$p_1, p_2, ..., p_N$ from truncated stick breaking mechanism(γ, H)
3. Generate N paths of the Markov Chain and place the generalized dirichlet distribution over the N paths.
4. Choose one of the N paths according to the corresponding p
5. Iterate over the following steps:

(a) Given a chosen path X_i:

   i. Sample from the conditionals of each component of θ

(b) Sample p=$p_1, p_2, ..., p_N$ from the conditional of p

(c) Update the conditional of X according to the conditional above

(d) Choose one of the paths according to the distribution of X

Chapter 4

Markov Modulated Mean reverting OU process

The method described in this section is inspired by the method in [Srikanth K. Iyer, 2009] but has not been previously developed to solve the infrence problem for MMOUs.

4.1 Model Definition

Definition 19. *Let $(X_t)_{t\in\{0,1,..,T\}}$ be a Discrete time Markov chain on finite space $\mathcal{S}$: = 1, 2, ..., M. The probability transition matrix of Markov Chain, denoted by $\mathbb{P}$, is given by*

$$\mathbb{P}_{ij} \geq 0 \text{ for all } i,j \in \mathcal{S} \text{ and } \mathbb{P}_{ij} = P(X_{t+1} = j | X_t = i) \tag{4.1}$$

We define our model as :

Definition 20. ***Regime Switching Ornstein–Uhlenbeck Model**: For $t \in \{0, 1, .., T\}$, let X_t be a discrete time Markov chain with finite state space $\mathcal{M} = \{1, \ldots, M\}$ defined as above. A regime-switching Ornstein–Uhlenbeck model is a stochastic process S_t which can be described by the equation below:*

$$dS_t \;=\; \kappa(X_t)(\theta(X_t) - S_t)dt + \sigma(X_t)dW, \tag{4.2}$$

where $\kappa(X_t), \theta(X_t), \sigma(X_t)$ are functions of the Markov chain X and W is a Brownian motion. Where $\kappa(X_t)$ and $\sigma(X_t) \in \mathbb{R}_+$, while $\theta(X_t) \in \mathbb{R}$

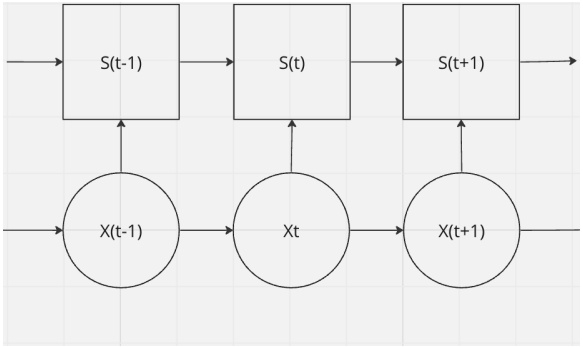


Fig. 4.1: X_t is the underlying Markov Chain while $S(t)$ is the MMOU process

Proposition 5. *If S_t is a stochastic process that is solution to the Stochastic differntial equation:*

$$dS_t = \kappa(\theta - S_t)dt + \sigma dW_t$$

Then the distribution of S_t can be given as

$$(S_t|S_0 = s_0) \sim \mathcal{N}(s_0 e^{-\kappa t} + (1 - e^{-\kappa t})\theta, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}))$$

Proof. Let $U_t = S_t - \theta$
then U_t satisfies the SDE:

$$dU_t = -\kappa U_t dt + \sigma dW_t$$

We again do a change of variable by multiplying by exponential of κt:

$$U_t = e^{-\kappa t} V_t$$

By Ito's integral, we can calculate that :

$$\begin{aligned} dV_t &= \kappa e^{\kappa t} U_t dt + e^{\kappa t} dU_t \\ &= \kappa e^{\kappa t} U_t dt + e^{\kappa t}(-\kappa U_t dt + \sigma dW_t) \\ &= \sigma e^{\kappa t} dW_t \end{aligned}$$

Now integrating both the sides and using Ito's rule we get:

$$V_t = V_s + \sigma \int_s^t e^{\kappa y} dW_y$$

now to get the original solution we change variables again.

$$U_t = e^{-\kappa t} V_t = e^{-\kappa(t-s)} U_s + \sigma e^{-\kappa t} \int_s^t e^{\kappa y} dW_y$$

$$S_t = U_t + \theta = \theta + e^{-\kappa(t-s)}(S_s - \theta) + \sigma \int_s^t e^{-\kappa(t-y)} dW_y$$

Now observe that the above solution has two parts. first an exponential function and second an integral with respect to Brownian motion.Therefore S would be normally distributed. So now we calculate the expectation and Variance of our process S.

$$E[S_t|S_0 = s_0] = E[\theta + e^{-\kappa(t)}(S_0 - \theta) + \sigma \int_0^t e^{-\kappa(t-y)} dW_y] \quad = \theta + (s_0 - \theta)e^{-\kappa t}$$

We find the variance as:

$$\begin{aligned} Var[S_t|S_0 = s_0] &= E[(\sigma \int_0^t e^{\kappa(t-y)} dW_y)^2] \\ &= \sigma^2 E[\int_0^t e^{-2\kappa(t-y)} dy)] \qquad \text{We use Itos Isometry} \\ &= \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}) \end{aligned}$$

Therefore we finally get :

$$(S_t|S_0 = s_0) \sim \mathcal{N}(s_0 e^{-\kappa t} + (1 - e^{-\kappa t})\theta, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}))$$

□

4.2 Inference Procedure

Now, we want to fit a regime-switching OU model such as above. Thus the optimal set of parameters to estimate is $(\hat{\theta}_i, \hat{\kappa}_i, \hat{\sigma}_i, \hat{X}_t) \in \mathbb{R} \times \mathbb{R}_+^{2\times M} \times \mathcal{D}[\{0, 1, ..., T\}, \{1, ..., M\}]$. where $i = 1, ..., M$.

Where $(\hat{\theta}_i, \hat{\kappa}_i, \hat{\sigma}_i)$ are the parameters for the OU process and $\hat{X(t)} \in \mathcal{D}[\{0, 1, ..., T\}, \{1, ..., M\}]$ is the path taken by the underlying Markov Chain and $\mathcal{D}[\{0, 1, ..., T\}, \{1, ..., M\}]$ is the set of possible paths for the Discrete time Markov Chain
We use the "Dirichlet prior method" as explained in the previous section. Therefore We first define the priors

4.2.1 Priors for the parameters

1. **Placing a Truncated Dirichlet prior over the Truncated Markov Chain Path Space**

   - We place a Truncated Dirichlet prior D(γ, H), on the truncated path space of the DTMC (X_t), with precision parameter γ and mean H which is a probability measure governing a DTMC on the path space $\mathcal{D}$, the set of cadlag functions.
   - We define distribution H as follows:
     (a) The initial distribution according to H is the uniform distribution π_0 = (1/M,..., 1/M)
     (b) The transition matrix according to distribution H is $\mathbb{P}$ with $\mathbb{P}_{ij}$ = 1/(M), $i, j = 1, ..., M$.
   - We then generate a N number of paths $\{X^i_{s\in\{0,...,T\}}, i = 1, ..., N\}$ from H.
   - Then we generate the vector of probabilities $\{p_i, i = 1, ..., N\}$ from a truncated stick-breaking scheme with parameter γ
   - Finally we can define the prior probability distribution over the truncated path space as:
   $$P(.) = \sum_{i=1}^{N} p_i \delta_{X^i}(.)$$

2. **Prior for discretization Parameter** γ We also place a Gamma prior for discretization Parameter γ.

   $$\gamma \sim \Gamma(\epsilon_1, \epsilon_2)$$

3. **Prior for the instantaneous volatility** σ

Since square of instantaneous volatility can vary over the positive real line we will use an inverse gamma prior for the instantaneous volatility.

$$\sigma^2 \sim \mathcal{IG}(\nu_1, \nu_2) \tag{4.3}$$

4. **Prior for long term mean level θ**
   Since the long term mean level can vary over the whole real space we place a normal prior over the mean level.

$$\theta \sim \mathcal{N}(\mu, \rho) \tag{4.4}$$

5. **prior for speed of reversion κ**
   We place an inverse gamma prior over the mean reversion rate as it also varies over the real line.

$$\kappa \sim \mathcal{IG}(\eta_1, \eta_2) \tag{4.5}$$

4.2.2 Gibbs Sampling Procedure

To carry out Gibbs Sampling we need to sample from the conditional for each of the parameter defined above. We therefore derive these conditionals below:

Proposition 6. *Conditional for σ^2*
The conditional for σ^2 is as follows. If we have a given MC path $X = \mathcal{X}$. Let $\mathcal{X}^*$ be the list of unique states taken by the MC path.

Then we will for each $j \in \mathcal{X}^*$ draw from:

$$(\sigma_j^2 | S, X, \kappa, \theta, S) \propto \prod_{k:\mathcal{X}_{t_{k-1}}=j} \{\frac{1}{b_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k}-a_j}{b_j})^2}\} \frac{\nu_2^{\nu_1}}{\Gamma(\nu_1)} (\sigma^2)^{(-\nu_1-1)} e^{-\frac{\nu_2}{\nu_1}} \tag{4.6}$$

Where $a_j = S_{t_{k-1}} e^{-\kappa(t_k - t_{k-1})} + (1 - e^{-\kappa(t_k - t_{k-1})})\theta$ *and* $b_j^2 = \frac{\sigma^2}{2\kappa}(1 - e^{-\kappa 2(t_k - t_{k-1})})$

Proof. From the Bayes theorem we have

$$
\begin{aligned}
&f(\sigma_j^2 = \text{II}|S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \\
&= \frac{f(\sigma_j^2 = \text{II}, S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(\sigma_j^2 = \int, S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) dP(\int)} \\
&= \frac{f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(S|\sigma_j^2 = \int, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \int X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) dP(\int)} \\
&= \frac{f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \text{II}|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \cancel{f(X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}}{\int_{\sigma_j^2} f(S|\sigma_j^2 = \int, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \int |X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \cancel{f(X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})} dP(\int)} \\
&= \frac{f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \text{II}|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(S|\sigma_j^2 = \int, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \int, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) dP(\int)} \\
&= \frac{f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \text{II}|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(\sigma_j^2 = \int, S|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) dP(\int)} \\
&= \frac{f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \text{II}|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}{f(S|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})}
\end{aligned}
$$

We can see the denominator is not dependent on σ_j^2 ,therefore we can write:

$$f(\sigma_j^2 = \text{II}|S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \propto f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) f(\sigma_j^2 = \text{II}|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T})$$

Now from 5 , we can observe

$$
\begin{aligned}
&f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \\
&= \prod_{k:\mathcal{X}_{t_{k-1}}=j} \mathcal{N}(S_{t_k}; S_{t_{k-1}} e^{-\mathcal{K}(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}(t_k - t_{k-1})})\mathcal{T}, \frac{\sigma^2}{2\mathcal{K}}(1 - e^{-\mathcal{K}2(t_k - t_{k-1})})
\end{aligned}
$$

Now, if we let $a_j = S_{t_{k-1}} e^{-\mathcal{K}(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}(t_k - t_{k-1})})\mathcal{T}$ and $b_j^2 = \frac{\sigma^2}{2\mathcal{K}}(1 - e^{-\mathcal{K}2(t_k - t_{k-1})})$
Then we can write

$$f(S|\sigma_j^2 = \text{II}, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) = \prod_{k:\mathcal{X}_{t_{k-1}}=j} \{\frac{1}{b_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - a_j}{b_j})^2}\}$$

Now, for the second term we know the prior of σ^2, therefore it would be as follows:

$$\begin{aligned} f(\sigma_j^2 = \amalg | X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) &= \mathcal{IG}(\sigma^2; \nu_1, \nu_2) \\ &= \frac{\nu_2^{\nu_1}}{\gamma(\nu_1)} (\sigma^2)^{-\nu_1 - 1} e^{-\frac{\nu_2}{\nu_1}} \end{aligned}$$

Therefore we finally get:

$$f(\sigma_j^2 | S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \propto \{ \prod_{k: \mathcal{X}_{t_{k-1}} = j} \frac{1}{b_j \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - a_j}{b_j})^2} \} \frac{\nu_2^{\nu_1}}{\Gamma(\nu_1)} (\sigma^2)^{(-\nu_1 - 1)} e^{-\frac{\nu_2}{\nu_1}}$$

□

Proposition 7. ***Conditional for*** κ

Given MC path $X = \mathcal{X}$ *and* $\mathcal{X}^*$ *be the list of unique states . For each* $j \in \mathcal{X}^*$ *draw :*

$$(\kappa_j | S, X = \mathcal{X}, \sigma_j^2 = \amalg, \theta_j = \mathcal{T}) \propto \{ \prod_{k: \mathcal{X}_{t_{k-1}} = j} \frac{1}{d_j \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - c_j}{d_j})^2} \} \frac{\eta_2^{\eta_1}}{\Gamma(\eta_1)} (\kappa)^{(-\eta_1 - 1)} e^{\frac{\eta_2}{\eta_1}} \quad (4.7)$$

Where $c_j = S_{t_{k-1}} e^{-\kappa_j (t_k - t_{k-1})} + (1 - e^{-\kappa_j (t_k - t_{k-1})}) \mathcal{T}$ *and* $d_j^2 = \frac{\amalg^2}{2\kappa} (1 - e^{-\kappa_j 2 (t_k - t_{k-1})})$

Proof. From the Bayes theorem we have

$$\begin{aligned} & f(\kappa = \mathcal{K} | S, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) \\ &= \frac{f(\kappa = \mathcal{K}, S, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(\kappa = \updownarrow, S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) dP(\updownarrow)} \\ &= \frac{f(S | \kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(S | \kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) dP(\updownarrow)} \\ &= \frac{f(S | \kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \mathcal{K} | X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \cancel{f(X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}}{\int_{\sigma_j^2} f(S | \kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \updownarrow | X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) \cancel{f(X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}} \\ &= \frac{f(S | \kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \mathcal{K} | X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(S | \kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) dP(\updownarrow)} \\ &= \frac{f(S | \kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \mathcal{K} | X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(\kappa = \updownarrow, S | X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) dP(\updownarrow)} \\ &= \frac{f(S | \kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \mathcal{K} | X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})}{f(S | X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})} \end{aligned}$$

We can see the denominator is not dependent on κ_j ,therefore we can write:

$$f(\kappa = \mathcal{K}|S, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) \propto f(S|\kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) f(\kappa = \mathcal{K}|X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})$$

Now from 5 , we can observe

$$\begin{aligned} & f(S|\kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) \\ & = \prod_{k:\mathcal{X}_{t_{k-1}}=j} \mathcal{N}(S_{t_k}; S_{t_{k-1}} e^{-\kappa(t_k - t_{k-1})} + (1 - e^{-\kappa(t_k - t_{k-1})})\mathcal{T}, \frac{\amalg^2}{2\kappa}(1 - e^{-\kappa 2(t_k - t_{k-1})}) \end{aligned}$$

Now, if we let $c_j = S_{t_{k-1}} e^{-\kappa(t_k - t_{k-1})} + (1 - e^{-\kappa(t_k - t_{k-1})})\mathcal{T}$ and $d_j^2 = \frac{\amalg^2}{2\kappa}(1 - e^{-\kappa 2(t_k - t_{k-1})})$
Then we can write

$$f(S|\sigma^2 = \amalg, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) = \prod_{k:\mathcal{X}_{t_{k-1}}=j} \{\frac{1}{d_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - c_j}{d_j})^2}\}$$

Now, for the second term we know the prior of κ, therefore it would be as follows:

$$\begin{aligned} f(_j = \mathcal{K}|X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) &= \mathcal{IG}(\kappa; \eta_1, \eta_2) \\ &= \frac{\eta_2^{\eta_1}}{\Gamma(\eta_1)} (\kappa)^{(-\eta_1 - 1)} e^{-\frac{\eta_2}{\eta_1}} \end{aligned}$$

Therefore we finally get:

$$f(\kappa|S, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T}) \propto \{\prod_{k:\mathcal{X}_{t_{k-1}}=j} \frac{1}{d_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - c_j}{d_j})^2}\} \frac{\eta_2^{\eta_1}}{\Gamma(\eta_1)} (\kappa)^{(-\eta_1 - 1)} e^{\frac{\eta_2}{\eta_1}}$$

□

Proposition 8. ***Conditional for*** θ
Given MC path $X = \mathcal{X}$ *and* $\mathcal{X}^*$ *be the list of unique states . For each* $j \in \mathcal{X}^*$ *draw :*

$$f(\theta_j|S, X = \mathcal{X}, \sigma_j^2 = \amalg, \kappa_j = \mathcal{K}) \propto \{\prod_{k:\mathcal{X}_{t_{k-1}}=j} \frac{1}{d_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - c_j}{d_j})^2}\} \frac{1}{\gamma_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\theta_j - \mu_j}{\gamma_j})^2} \quad (4.8)$$

Where $c_j = S_{t_{k-1}} e^{-\mathcal{K}(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}(t_k - t_{k-1})})\theta$ *and* $d_j^2 = \frac{\amalg^2}{2\mathcal{K}}(1 - e^{-\mathcal{K}2(t_k - t_{k-1})})$

Proof. From the Bayes theorem we have

$$
\begin{aligned}
&f(\theta_j = \mathcal{T}|S, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K}) \\
&= \frac{f(\theta = \mathcal{T}, S, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})}{\int_{\theta_j} f(\theta = \sqcap, S, X = \mathcal{X}, \sigma^2 = \amalg\theta = \mathcal{T})dP(\sqcap)} \\
&= \frac{f(S|\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})f(\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})}{\int_{\theta_j} f(S|\kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \sqcap)f(\kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \sqcap)dP(\sqcap)} \\
&= \frac{f(S|\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})f(\theta = \mathcal{T}|X = \mathcal{X}, \kappa = \mathcal{K}, \sigma^2 = \amalg)f(X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})}{\int_{\theta_j} f(S|\kappa = \updownarrow, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \sqcap)f(\theta = \sqcap|X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})f(X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})} \\
&= \frac{f(S|\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})f(\theta = \mathcal{T}|X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})}{\int_{\theta_j} f(S|\theta = \sqcap, \kappa = \mathcal{K}, X = \mathcal{X}, \sigma^2 = \amalg)f(\theta = \sqcap, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})dP(\sqcap)} \\
&= \frac{f(S|\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \theta = \mathcal{T})f(\theta = \mathcal{T}|X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})}{\int_{\theta_j} f(\theta = \sqcap, S|X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})dP(\sqcap)} \\
&= \frac{f(S|\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})f(\theta = \mathcal{T}|X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})}{f(S|X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K}}
\end{aligned}
$$

We can see the denominator is not dependent on θ_j ,therefore we can write:

$$f(\theta = \mathcal{T}|S, X = \mathcal{X}, \kappa = \mathcal{K}, \sigma^2 = \amalg) \propto f(S|\theta = \mathcal{T}, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K})f(\theta = \mathcal{T}|X = \mathcal{X}, \sigma^2 = \amalg,$$

Now from 5 , we can observe

$$
\begin{aligned}
&f(S|\theta, X = \mathcal{X}, \sigma^2 = \amalg, \kappa = \mathcal{K}) \\
&= \prod_{k:\mathcal{X}_{t_{k-1}}=j} \mathcal{N}(S_{t_k}; S_{t_{k-1}}e^{-\mathcal{K}(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}(t_k - t_{k-1})})\theta, \frac{\amalg^2}{2\kappa}(1 - e^{-\mathcal{K}2(t_k - t_{k-1})})
\end{aligned}
$$

Now, if we let $c_j = S_{t_{k-1}}e^{-\mathcal{K}(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}(t_k - t_{k-1})})\theta$ and $d_j^2 = \frac{\amalg^2}{2\mathcal{K}}(1 - e^{-\mathcal{K}2(t_k - t_{k-1})})$
Then we can write

$$f(S|\theta, \sigma^2 = \amalg, X = \mathcal{X}, \kappa = \mathcal{K},) = \prod_{k:\mathcal{X}_{t_{k-1}}=j} \{\frac{1}{d_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - c_j}{d_j})^2}\}$$

Now, for the second term we know the prior of θ, therefore it would be as follows:

$$\begin{aligned} f(\theta_j = \mathcal{K}|X = \mathcal{X}, \kappa = \mathcal{K}, \sigma^2 = \Pi) &= \mathcal{N}(\theta; \mu_j, \gamma_j) \\ &= \frac{1}{\gamma_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\theta-\mu_j}{\gamma_j})^2} \end{aligned}$$

Therefore we finally get:

$$f(\theta|S, X = \mathcal{X}, \sigma^2 = \Pi, \theta = \mathcal{T}) \propto \{ \prod_{k:\mathcal{X}_{t_{k-1}}=j} \frac{1}{d_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k}-c_j}{d_j})^2} \} \frac{1}{\gamma_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\theta-\mu_j}{\gamma_j})^2}$$

□

Proposition 9. *conditional for H*
The re-estimation of the initial distribution and the transition matrix from a given path Xi can be done by any standard MLE procedure for CTMC. We have chosen to use simple gradient descent method for our algorithm. We also use the analytical solution for the maximum likelihood estimate for the ctmc.

Proposition 10. *conditional for X*

$$f(X = x|p, \sigma, \kappa, \theta) \sim \sum_{i=1}^{N} p_i^* \delta_{X_i} \tag{4.9}$$

$$p_i^* \propto \prod_{j=1}^{m} \{ \prod_{k:x^{i,*}_{t^i_{k-1}}=j} \frac{1}{b_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k}-a_j}{b_j})} \} p_i \tag{4.10}$$

where $X_i = (x_1^i, ..., x_n^i)$ and $(x_1^{i,*}, ..., x_n^{i,*})$ denote the current m unique values in the path X_i

Proof. We know from the use of bayes theorem that

$$f(X = x|p, W, \sigma, \kappa, \theta)$$

$$= \frac{f(S|\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})f(\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_X f(S|\kappa = \updownarrow, X = y, \sigma^2 = \amalg, \theta = \mathcal{T})f(\kappa = \updownarrow, X = y, \sigma^2 = \amalg, \theta = \mathcal{T})dP(y)}$$

$$= \frac{f(S|\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})f(X = x|\sigma^2 = \amalg, \kappa = \mathcal{K}, \theta = \mathcal{T})\cancel{f(X = x, \sigma^2 = \amalg, \theta = \mathcal{T})}}{\int_X f(S|\kappa = \updownarrow, X = y, \sigma^2 = \amalg, \theta = \mathcal{T})f(X = y|\kappa = \updownarrow, \sigma^2 = \amalg, \theta = \mathcal{T})\cancel{f(\kappa = \updownarrow, \sigma^2 = \amalg, \theta = \mathcal{T})}dP}$$

$$= \frac{f(S|\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})f(X = x|\kappa = \mathcal{K}, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_X f(S|\kappa = \updownarrow, X = y, \sigma^2 = \amalg, \theta = \mathcal{T})f(\kappa = \updownarrow, X = y, \sigma^2 = \amalg, \theta = \mathcal{T})dP(y)}$$

$$= \frac{f(S|\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})f(X = x|\kappa = \mathcal{K}, \sigma^2 = \amalg, \theta = \mathcal{T})}{\int_{\sigma_j^2} f(X = x, S|\kappa = \updownarrow, \sigma^2 = \amalg, \theta = \mathcal{T})dP(y)}$$

$$= \frac{f(S|\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})f(X = x|\kappa = \mathcal{K}, \sigma^2 = \amalg, \theta = \mathcal{T})}{f(S|\kappa = \mathcal{K}, \sigma^2 = \amalg, \theta = \mathcal{T})}$$

We can see the denominator is not dependent on X,therefore we can write:

$$f(X = x|p, W, \sigma, \kappa, \theta) \propto f(S|\kappa = \mathcal{K}, X = x, \sigma^2 = \amalg, \theta = \mathcal{T})f(X = x|\kappa = \mathcal{K}, \sigma^2 = \amalg, \theta = \mathcal{T})$$

Now from 5 , we can observe

$$f(S|\sigma^2 = \amalg, X = x, \kappa = \mathcal{K}, \theta = \mathcal{T})$$
$$= \prod_{j=1}^{m} \prod_{k:x_{t_{k-1}}=j} \mathcal{N}(S_{t_k}; S_{t_{k-1}}e^{-\mathcal{K}_j(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}_j(t_k - t_{k-1})})\mathcal{T}_j, \frac{\amalg_j^2}{2\mathcal{K}_j}(1 - e^{-\mathcal{K}_j 2(t_k - t_{k-1})})$$

Now, if we let $a_j = S_{t_{k-1}}e^{-\mathcal{K}_j(t_k - t_{k-1})} + (1 - e^{-\mathcal{K}_j(t_k - t_{k-1})})\mathcal{T}_j$ and $b_j^2 = \frac{\amalg_j}{2\mathcal{K}}(1 - e^{-\mathcal{K}_j 2(t_k - t_{k-1})})$
Then we can write

$$f(S|\sigma_j^2 = \amalg, X = \mathcal{X}, \kappa = \mathcal{K}, \theta = \mathcal{T}) = \prod_{j=1}^{m} \prod_{k:x_{t_{k-1}}=j} \{\frac{1}{b_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - a_j}{b_j})^2}\}$$

Also,$f(X = x|p, \sigma, \kappa, \theta) = f(X|p) = \sum_{k=1}^{N} p_i \delta_{X_i}(.)$ Using the above facts we can write

the following.

$$f(X = x|p, \sigma, \kappa, \theta) \propto f(S|X, p, \sigma, \kappa, \theta) \sum_{k=1}^{N} p_i \delta_{X_i}(x)$$
$$\propto \sum_{i=1}^{N} \delta_{X_i}(x) p_i \{ \prod_{j=1}^{m} \{ \prod_{k: x_{t_{k-1}}^{i,*} = j} \frac{1}{b_j \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k} - a_j}{b_j})} \} \}$$

□

Proposition 11. *conditional for* $P = (p_1, ..., p_N)$

$$p_1 = V_1^* \quad and \quad p_k = (1 - V_1^*)...(1 - V_{k-1}^*)V_k^* \tag{4.11}$$

where

$$(V_k^*|X_j) \sim \beta(a_k^*, b_k^*) \tag{4.12}$$

$$a_k^* = \begin{cases} a_k^* = a_k + 1 \quad if j = k \\ a_k^* = a_k \quad else if j \neq k \end{cases}$$
$$b_k^* = \begin{cases} b_k^* = b_k + 1 \quad if j > k > 0 \\ b_k^* = b_k \quad else if j \leq k \end{cases}$$

Proof. We first note that according to our model $P(X = j|P) = p_j$.
Therefore

$$f(P|X = j) = \frac{f(X = j|P) * f(P)}{\int_p f(X = j|P) * f(P) dP} \qquad \text{Using the Bayes theorem}$$
$$\propto f(X = j|P) * f(P)$$
$$\propto p_j * \{ \prod_{i=1}^{N-1} \mathcal{B}(a_i, b_i)] p_N^{b_{N-1}-1} \prod_{i=1}^{N-1} [p_i^{a_i - 1} (\sum_{k=i}^{N} p_k)^{b_{i-1} - (a_i + b_i)} \} \qquad \text{Using Theorem 1}$$

Let $P_i = \sum_{k=i}^{N} p_k$. then we can say.

$$f(P|X=j) \propto \{p_N^{b_{N-1}-1}\prod_{i=1}^{j-1} p_i^{a_i-1}(P_i)^{(b_{i-1}+1)-(a_i+(b_i+1))}\}p_j^{(a_j+1)-1}P_j^{(b_{j-1}+1)-((a_j+1)+(b_j))}\}$$
$$\{\prod_{i=1}^{N-1} p_i^{a_i-1}(P_i)^{b_{i-1}-(a_i+b_i)}\} \quad (4.13)$$

From the above we can deduce that the vector P is distributed $\mathcal{G}(a_1^*, b_1^*, ..., a_N^*, b_N^*)$. Which implies

$$p_1 = V_1^* \quad and \quad p_k = (1-V_1^*)...(1-V_{k-1}^*)V_k^* \quad (4.14)$$

where

$$(V_k^*|X_j) \sim \beta(a_k^*, b_k^*) \quad (4.15)$$

$$a_k^* = \begin{cases} a_k^* = a_k + 1 \quad if j = k \\ a_k^* = a_k \quad else if j \neq k \end{cases}$$
$$b_k^* = \begin{cases} b_k^* = b_k + 1 \quad if j > k > 0 \\ b_k^* = b_k \quad else if j \leq k \end{cases}$$

□

Proposition 12. *conditional for γ is*

$$f(\gamma|p) = \Gamma(N + \epsilon_1 - 1, \epsilon_2 - \sum_{i=1}^{N-1} \log(1 - V_i^*)) \quad (4.16)$$

Proof. We can observe from Theorem 1 that the distribution for P is

$$\{\prod_{i=1}^{N-1} \mathcal{B}(a_i, b_i)]p_N^{b_{N-1}-1}\prod_{i=1}^{N-1}[p_i^{a_i-1}(\sum_{k=i}^{N} p_k)^{b_{i-1}-(a_i+b_i)}\} \quad (4.17)$$

so now using the fact that we have $a_i = 1$and $b_i = \gamma$ and also using the fact that

$\Gamma(1+\gamma) = \gamma\Gamma(\gamma)$ we get that

$$f(p|\gamma) \propto \gamma^{N-1} p_N^{\gamma-1} = \gamma^{N-1} e^{(\gamma-1)\log(p_N)} \tag{4.18}$$

Therefore we can finally write using Bayes theorem

$$\begin{aligned}
f(\gamma|p) &\propto f(p|\gamma) f(\gamma) \\
&\propto \gamma^{N-1} e^{(\gamma-1)\log(p_N)} * \{\gamma^{\epsilon_1 - 1} e^{-\epsilon_2 \gamma}\} && \text{from the Gamma distribution} \\
&\propto \gamma^{N+\epsilon_1-2} e^{-(\epsilon_2 - \log(p_N))\gamma} e^{-\log(p_N)} && \text{Its an un-normalized Gamma distribution} \\
f(\gamma|p) &= \Gamma(N + \epsilon_1 - 1, \epsilon_2 - \log(p_N)) && \{p_N = \prod_{i=1}^{N-1} V_i^*\} \\
f(\gamma|p) &= \Gamma(N + \epsilon_1 - 1, \epsilon_2 - \sum_{i=1}^{N-1} \log(1 - V_i^*))
\end{aligned}$$

□

4.3 Gibbs Sampler Algorithm

We give the "Truncated Dirichlet Prior Method algorithm for MMOU" as follows.

Algorithm 1: Truncated DP Algorithm for Mean Reverting OU process

Choose hyper parameters ϵ_1, ϵ_2 ;
Generate γ from Gamma(ϵ_1, ϵ_2)
Generate N paths of the Markov Chain from the distribution H.
Draw (p=$p_1, p_2, ..., p_N$) from stick breaking with mechanism(γ, H)
Choose one of the N paths according to the corresponding p
for $i = 1$ *to* . **do**
    Update the conditional of X according to the conditional above
    Choose one of the paths according to the distribution of X;
    Given a path X update the conditionals for all the parameters:
        Draw a new σ^2 according to the equation 4.6
        Draw a new κ according to the equation4.7
        Draw a new θ according to the equation4.8
    Define a_k^* and b_k^* according to equation 4.12 where k is the index of chosen path
    Compute $p_i's$ according to the equation 4.11
    Draw a new γ according to the equation 4.16
end
Draw the most probable Markov Chain Path from the posterior distribution.
Estimate H according to the Maximum Likelihood Estimation of the most probable Markov Chain Path.

Chapter 5

Markov Modulated Geometric Levy process

The method described in this section is inspired by the method in [Srikanth K. Iyer, 2009] but has not been previously developed to solve the infrence problem for MMLPs.

5.1 Model Definition

Definition 21. *Let $(X_t)_{t\in[0,T]}$ be a continuous time Markov chain on finite space S: = 1, 2, ..., M. The generator matrix of M, denoted by $\mathbb{Q}$, is given by*

$$Q_{ij} \geq 0 \text{ if } i \neq j \text{ for all } i,j \in S \text{ and } Q_{ii} = -\sum_{j\neq i} Q_{ij} \text{ otherwise} \tag{5.1}$$

We define our model as :

Definition 22. *Regime Switching Geometric Levy Model: For all t∈[0, T], let S_t be a continuous time Markov chain on finite space $\mathcal{S} = \{1, \ldots, M\}$ defined as above. A Regime Switching Geometric Levy model is a stochastic process S_t which can be described the equation below:*

$$S_t \quad = \quad \exp Y_t^{\zeta}(X_t) \tag{5.2}$$

where Y_t^{ζ} is a regime switching NIG Levy process.

Proposition 13. *If S_t is a stochastic process that can be described according to the*

following defintion:

$$S_t = \exp Y_t$$

Then the distribution of $Z_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$ *can be given as*

$$Z_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \sim NIG(\alpha, \beta, \delta, \mu)$$

Proof. We have first

$$\begin{aligned} ln(S_t) &= Y_t && \text{Since } Y_t \text{ is NIG process} \\ ln(S_t) - ln(S_{t-1}) &= Y_t - Y_{t-1} \sim NIG(\alpha, \beta, \delta, \mu) && \text{Finally} \\ Z_t &= \ln\left(\frac{S_t}{S_{t-1}}\right) \sim NIG(\alpha, \beta, \delta, \mu) \end{aligned}$$

□

Proposition 14. *If A and B are random variables such that, A*$\sim NIG(\alpha, \beta, \mu_1, \delta_1)$ *and B*$\sim NIG(\alpha, \beta, \mu_2, \delta_2)$ *. Then*

$$A + B \sim NIG(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$$

Proposition 15. *Given the path* $X = \{X_s, 0 \leq s \leq n\}$*, let* $T_j(t)$ *be the time spent by the path X in state j in the time interval [t - 1, t] Define*

$$\mu(t) = \sum_{j=1}^{M} \mu_j T_j(t) \tag{5.3}$$

$$\delta(t) = \sum_{j=1}^{M} \delta_j T_j(t) \tag{5.4}$$

Then we have conditional on the path X, $Z_t \sim NIG(\alpha, \beta, \mu(t), \delta(t))$*, t=1,2,....,n .*

5.2 Inference Procedure

Now, we want to fit a regime-switching exponential levy model such as above. Thus the optimal set of parameters to estimate is $(\hat{\mu}_i, \hat{\delta}_i, \hat{\alpha}, \hat{\beta}, \hat{X(t)}) \in \mathbb{R}^{2\times M} \times \mathbb{R}^2 \times \mathcal{D}$.

where $i = 1, ..., M$.
Where $(\hat{\mu}_i, \hat{\delta}_i, \hat{\alpha}, \hat{\beta})$ are the parameters for the NIG process if $X(t) = i$ for all $i = 1, ..., M$ and $\hat{X(t)} \in \mathcal{D}$ is the path taken by the underlying Markov Chain and $\mathcal{D}$ is the set of Cadlag paths of the continuous time Markov Chain
We use the "Dirichlet prior method" as explained in the previous section. Therefore We first define the priors

5.2.1 Priors for the parameters

1. **Placing a Dirichlet prior over the Markov Chain**

   - We place a Truncated Dirichlet prior D(γ, H), on the truncated path space of the CTMC (X_t), with precision parameter γ and mean H which is a probability measure for the CTMC path space $\mathcal{D}$, the set of cadlag functions.
   - We define distribution H as follows:
     (a) H defines the initial distribution as the uniform distribution. π_0 = (1/M,..., 1/M)
     (b) The Q matrix according to distribution H is $\mathbb{Q}$ with $\mathbb{Q}_{ij} = 1/(\text{M-1})$, for all i,j s.t. $i \neq j$.
   - We then generate a N number of paths $\{X^i_{s\in[0,T]}, i = 1, ..., N\}$ from H.
   - Then we generate the vector of probabilities $\{p_i, i = 1, ..., N\}$ from a truncated stick-breaking scheme with parameter γ
   - Finally we can define the prior probability distribution over the truncated path space as:
     $$P(.) = \sum_{i=1}^{N} p_i \delta_{X^i}(.)$$

2. **Prior for discretization Parameter** γ We also place a Gamma prior for discretization Parameter γ.

   $$\gamma \sim \Gamma(\epsilon_1, \epsilon_2)$$

For each i=1,2,...,M , we define the priors as follows

$$\mu_i \sim \mathcal{N}(\theta, \tau^\mu) \tag{5.5}$$

$$\delta_i \sim \textbf{HalfNormal}(\nu_1) \tag{5.6}$$

where **HalfNormal**(ν_1) denotes $f(x;\nu_1) = \frac{\sqrt{2}}{\nu_1\sqrt{\pi}} \exp\left(-\frac{x^2}{2\nu_1^2}\right), \quad x > 0$ with scale parameter ν_1.

We also place priors on the remaining parameters α and β as follows

$$\alpha \sim \mathcal{IG}(\alpha_1, \alpha_2) \tag{5.7}$$

$$\beta \sim \mathcal{IG}(\beta_1, \beta_2) \tag{5.8}$$

5.2.2 Preparing the observed data

To determine the conditional distribution of parameters,we first extract the change in log-returns between Markov chain jump times. Let $0 = t_0 < t_1 < t_2 < ... < t_J$ be the times at which the path X changes state. Then we define the log-returns as $W_k = log(S_{t_k}/S_{t_{k-1}})$, k = 1, 2, .. , J. To derive realizations of the W_k from the observed Z process, we must simulate NIG random variables conditioned on their sums.

Let t $\in\{0, 1, .. , n\}$ for which the Markov chain switches state minimum once in the time interval [t-1, t]. So for p, k we have $t_{k-1} < t-1 \leq t_k < ... < t_{k+p} \leq t < t_{k+p+1}$. Let $s_0 = t_k - (t-1), \quad s_i = t_{k+i} - t_{k+i-1}, \quad s_{p+1} = t - t_{k+p}$, then $m_j = \mu_j s_i$ and $v_j = \delta_j s_i$, i=0,1,..p+1.

Then we have the joint density of $(W_0, W_1, ..., W_{k+p})$, given $Z_t = z$ as follows

$$f(u_0, u_1, .., u_p) = \frac{\prod_{i=0}^{p} NIG(u_i;\alpha,\beta,\mu_i,\delta_i) NIG(z - \sum u_i;\alpha,\beta,\mu_{p+1},\delta_{p+1})}{NIG(z;\alpha,\beta,\mu_t,\delta_t)} \tag{5.9}$$

5.2.3 Gibbs Sampling procedure

We are now ready to estimate the posterior distributions of the parameters using Gibbs sampling. In the following secton we derive the conditional distribution for each of the parameters for the markov modulate levy process

X^* is the set of distinct values observed in the path X of the CTMC.

Proposition 16. *Conditional for μ*
The conditional for μ is as follows. for each $j \in X^*$ draw from

$$(\mu_j|\alpha,\beta_j,\delta,X,W) \propto \{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\} * \mathcal{N}(\mu_j;\theta,\tau^\mu) \tag{5.10}$$

similarly for each j$\in X - X^*$ independently draw from the $\mu_j \sim \mathcal{N}(\mu_j;\theta,\tau^\mu)$

Proof. Let $k_1, k_2, k_3, ...$ be the times at which the Markov chain takes value j.
Then using Bayes theorem and observing the fact that the state Markov chain will remain in state j till $k_i + 1$ which means W_{k_i+1} will have parameters according to state j we can state that.
Let $\kappa = (\alpha, \beta, \delta, X)$

$$f((\mu_j|\alpha,\beta,\delta,X,W) = \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)*\} * f(\mu_j|\kappa)f(\kappa)}{\int_{\mu_j}\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * f(\mu_j|\kappa)f(\kappa)d\mu_j}$$

And since μ_j doesn't depends on κ we can write the above equation in following way

$$\begin{aligned} f((\mu_j|\theta,\tau^\mu,\alpha,\beta,\delta,X,W) &= \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * f(\mu_j)}{\int_{\mu_j^*}\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * f(\mu_j^*)d\mu_j^*} \\ &= \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathcal{N}(\mu_j;\theta,\tau^\mu)}{\int_{\mu_j^*}\{\prod_{i=1}^n f_{NIG}(W_{k_i+1},\mu_j^*|\delta_j,\alpha,\beta,X)\} * d\mu_j^*} \\ &= \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathcal{N}(\mu_j;\theta,\tau^\mu)}{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1},|\delta_j,\alpha,\beta,X)\}} \end{aligned}$$

now since the denominator doesn't depend on μ_j, we can state

$$\begin{aligned} &\propto \prod_{i=1}^n \{f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathcal{N}(\mu_j;\theta,\tau^\mu) \\ &\propto \{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\} * \mathcal{N}(\mu_j;\theta,\tau^\mu) \end{aligned}$$

□

Proposition 17. *Conditional for δ*
The conditional for δ is as follows. The sampling can be done in two steps first for the observed states and then for the unobserved states as follows.

for each $j \in X^*$ *draw*

$$(\delta_j|\alpha,\beta,\mu,X,W) \sim \{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\delta_j,\mu_j,\alpha,\beta)\} * \mathbf{HalfNormal}(\delta_j;\nu_1) \quad (5.11)$$

similarly for each $j \in X - X^*$ *independently draw from the* $\delta_j \sim \mathbf{HalfNormal}(\delta_j;\nu_1)$

Proof. Let $k_1, k_2, k_3, ...$ be the times at which the Markov chain takes value j.
Then using Bayes theorem and observing the fact that the state Markov chain will remain in state j till $k_i + 1$ which means W_{k_i+1} will have parameters according to state j we can state that. Let $\kappa = (\alpha, \beta, \mu_j, X)$

$$\begin{aligned} f((\delta_j|\alpha,\beta,\delta,X,W) &= \frac{\{f_{NIG}(W_{k_1+1},W_{k_2+1},..|\mu_j,\delta_j,\alpha,\beta,X)\} * f(\delta_j|\kappa)f(\kappa)}{\int_{\delta_j^*}\{f_{NIG}(W_{k_1+1},W_{k_2+1},..|\delta_j^*,\mu_j,\alpha,\beta,X)\} * f(\delta_j^*|\kappa)f(\kappa)d\delta_j^*} \\ &= \frac{\{f_{NIG}(W_{k_1+1},W_{k_2+1},..|\mu_j,\delta_j,\alpha,\beta,X)\} * f(\delta_j|\kappa)\cancel{f(\kappa)}}{\int_{\delta_j^*}\{f_{NIG}(W_{k_1+1},W_{k_2+1},..|\delta_j^*,\mu_j,\alpha,\beta,X)\} * f(\delta_j^*|\kappa)\cancel{f(\kappa)}d\delta_j^*} \end{aligned}$$

And since δ_j doesn't depend on κ. We can write

$$\begin{aligned} f((\delta_j|\alpha,\beta,\delta,X,W) &= \frac{\{f_{NIG}(W_{k_1+1},W_{k_2+1},..|\mu_j,\delta_j,\alpha,\beta,X)\} * f(\delta_j)}{\int_{\delta_j^*}\{f_{NIG}(W_{k_1+1},W_{k_2+1},..|\delta_j^*,\mu_j,\alpha,\beta,X)\} * f(\delta_j^*)d\delta_j^*} \\ &= \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathbf{HalfNormal}(\delta_j;\nu_1)}{\int_{\delta_j^*}\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\delta_j^*,\mu_j,\alpha,\beta,X)\} * f(\delta_j^*)d\delta_j^*} \\ &= \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathbf{HalfNormal}(\delta_j;\nu_1)}{\int_{\delta_j^*}\{\prod_{i=1}^n f_{NIG}(W_{k_i+1},\delta_j^*|\mu_j,\alpha,\beta,X)\}d\delta_j^*} \\ &= \frac{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathbf{HalfNormal}(\delta_j;\nu_1)}{\{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\alpha,\beta,X)\}} \end{aligned}$$

Since the denominator doesn't depend on δ_j we can write the following

$$\begin{aligned} f((\delta_j|\alpha,\beta,\delta,X,W) &\propto \{\prod_{i=1}^n f_{NIG}(W_{k_i+1}|\mu_j,\delta_j,\alpha,\beta,X)\} * \mathbf{HalfNormal}(\delta_j;\nu_1) \\ &\propto \{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\} * \mathbf{HalfNormal}(\delta_j;\nu_1) \end{aligned}$$

□

Proposition 18. ***Conditional for*** α
Given a MC path X *and* X^* *the list of unique states in* X.

The conditional for α is as follows:

$$(\alpha|\mu,\beta,\delta,X,W) \propto \{\prod_{j\in X^*}\{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\}\} * \mathcal{IG}(\alpha;\alpha_1,\alpha_2) \quad (5.12)$$

Proof. Let $\kappa = (\mu,\beta,\delta,X)$
Then using Bayes theorem we can state the following:

$$\begin{aligned} f((\alpha|\mu,\beta,\delta,X,W) &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * f(\alpha|\kappa)f(\kappa)}{\int_{\alpha^*}\{f_{NIG}(W_1,W_1,..|\delta,\mu,\alpha^*,\beta,X)\} * f(\alpha^*|\kappa)f(\kappa)d\alpha^*} \\ &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * f(\alpha|\kappa)\cancel{f(\kappa)}}{\int_{\alpha^*}\{f_{NIG}(W_1,W_1,..|\delta,\mu,\alpha^*,\beta,X)\} * f(\alpha^*|\kappa)\cancel{f(\kappa)}d\alpha^*} \end{aligned}$$

And since α doesn't depend on κ. We can write

$$\begin{aligned} f((\alpha|\mu,\beta,\delta,X,W) &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * f(\alpha)}{\int_{\alpha^*}\{f_{NIG}(W_1,W_1,..|\delta,\mu,\alpha^*,\beta,X)\} * f(\alpha^*)d\alpha^*} \\ &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * \mathcal{IG}(\alpha;\alpha_1,\alpha_2)}{\int_{\alpha^*}\{f_{NIG}(W_1,W_1,..,\alpha^*|\delta,\mu,\beta,X)\}d\alpha^*} \\ &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * \mathcal{IG}(\alpha;\alpha_1,\alpha_2)}{\{f_{NIG}(W_1,W_1,..|\delta,\mu,\beta,X)\}} \end{aligned}$$

Since the denominator doesn't depend on α we can write the following

$$f((\alpha|\mu,\beta,\delta,X,W) \propto \{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * \mathcal{IG}(\alpha;\alpha_1,\alpha_2)$$

Let $k_1,k_2,k_3,...$ be the times at which the Markov chain takes value j and observe the fact that the state Markov chain will remain in state j till k_i+1 which means W_{k_i+1} will have parameters according to state j. Then can state that.

$$f((\alpha|\mu,\beta,\delta,X,W) \propto \{\prod_{j\in X^*}\{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\}\} * \mathcal{IG}(\alpha;\alpha_1,\alpha_2)$$

□

Proposition 19. ***Conditional for*** β
Given a MC path X and X^* the list of unique states in X.

The conditional for β is as follows:

$$(\beta|\mu,\alpha,\delta,X,W) \propto \{\prod_{j\in X^*}\{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\}\} * \mathcal{IG}(\beta;\beta_1,\beta_2) \quad (5.13)$$

Proof. Let $\kappa = (\mu,\beta,\alpha,X)$
Then using Bayes theorem we can state the following:

$$\begin{aligned} f((\beta|\mu,\alpha,\delta,X,W) &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\beta,\alpha,X)\} * f(\beta|\kappa)f(\kappa)}{\int_{\beta^*}\{f_{NIG}(W_1,W_1,..|\delta,\mu,\beta^*,\alpha,X)\} * f(\beta^*|\kappa)f(\kappa)d\beta^*} \\ &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\beta,\alpha,X)\} * f(\beta|\kappa)\cancel{f(\kappa)}}{\int_{\beta^*}\{f_{NIG}(W_1,W_1,..|\delta,\mu,\beta^*,\alpha,X)\} * f(\beta^*|\kappa)\cancel{f(\kappa)}d\beta^*} \end{aligned}$$

And since β doesn't depend on κ. We can write

$$\begin{aligned} f((\beta|\mu,\alpha,\delta,X,W) &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * f(\beta)}{\int_{\beta^*}\{f_{NIG}(W_1,W_1,..|\delta,\mu,\beta^*,\alpha,X)\} * f(\beta^*)d\beta^*} \\ &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\beta,\alpha,X)\} * \mathcal{IG}(\beta;\beta_1,\beta_2)}{\int_{\beta^*}\{f_{NIG}(W_1,W_1,..,\beta^*|\delta,\mu,\alpha,X)\}d\beta^*} \\ &= \frac{\{f_{NIG}(W_1,W_2,..|\mu,\delta,\beta,\alpha,X)\} * \mathcal{IG}(\beta;\beta_1,\beta_2)}{\{f_{NIG}(W_1,W_1,..|\delta,\mu,\alpha,X)\}} \end{aligned}$$

Since the denominator doesn't depend on β we can write the following

$$f((\beta|\mu,\alpha,\delta,X,W) \propto \{f_{NIG}(W_1,W_2,..|\mu,\delta,\alpha,\beta,X)\} * \mathcal{IG}(\beta;\beta_1,\beta_2)$$

Let $k_1,k_2,k_3,...$ be the times at which the Markov chain takes value j and observe the fact that the state Markov chain will remain in state j till k_i+1 which means W_{k_i+1} will have parameters according to state j. Then can state that.

$$f((\beta|\mu,\alpha,\delta,X,W) \propto \{\prod_{j\in X^*}\{\prod_{k:X_{t_k-1}=j} f_{NIG}(W_k|\mu_j,\delta_j,\alpha,\beta)\}\} * \mathcal{IG}(\beta;\beta_1,\beta_2)$$

□

Proposition 20. *conditional for H*
The re-estimation of the initial distribution and the transition matrix from a given path Xi can be done by any standard MLE procedure for CTMC. We have chosen to use simple gradient descent method for our algorithm. We also use the analytical

solution for the maximum likelihood estimate for the ctmc.

Proposition 21. *conditional for X*

$$f(X = x|p, \mu, \delta, \alpha, \beta) \sim \sum_{i=1}^{N} p_i^* \delta_{X_i}$$

$$p_i^* \propto \{\prod_{j=1}^{m}\{\prod_{k:x^{i,*}_{t^i_{k-1}}=j} f_{NIG}(W_k|\mu_j, \delta_j, \alpha, \beta)\}\}p_i \tag{5.14}$$

where $X_i = (x_1^i, ..., x_n^i)$ *and* $(x_1^{i,*}, ..., x_n^{i,*})$ *denote the current m unique values in the path* X_i

Proof. We know from the use of bayes theorem that

$$f(X = x|p, W, \mu, \delta, \alpha, \beta) = \frac{f(X = x, p, W, \mu, \delta, \alpha, \beta)}{\int_X f(X = y, p, W, \mu, \delta, \alpha, \beta)dP(y)}$$

For the numerator we have,

$$f(X = x, p, W, \mu, \delta, \alpha, \beta) = f(W|X, p, \mu, \delta, \alpha, \beta)f(X = x|p, \mu, \delta, \alpha, \beta)f(p, \mu, \delta, \alpha, \beta)$$

Since the distribution of $(p, \mu, \delta, \alpha, \beta)$ doesn't depend on X we can cancel it out from the numerator as well as the denominator. Also,$f(X = x|p, \mu, \delta, \alpha, \beta) = f(X|p) = \sum_{k=1}^{N} p_i \delta_{X_i}(.)$ Using the above facts we can write the following.

$$f(X = x|p, \mu, \delta, \alpha, \beta) \propto f(W|X, p, \mu, \delta, \alpha, \beta) \sum_{k=1}^{N} p_i \delta_{X_i}(x)$$

$$\propto \sum_{i=1}^{N} \delta_{X_i}(x)\{\prod_{j=1}^{m}\{\prod_{k:x^{i,*}_{t^i_{k-1}}=j} \frac{1}{b_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{S_{t_k}-a_j}{b_j})}\}p_i\}$$

□

Proposition 22. *conditional for* $P = (p_1, ..., p_N)$

$$p_1 = V_1^* \quad and \quad p_k = (1 - V_1^*)...(1 - V_{k-1}^*)V_k^* \tag{5.15}$$

where

$$(V_k^*|X_j) \sim \beta(a_k^*, b_k^*) \tag{5.16}$$

$$a_k^* = \begin{cases} a_k^* = a_k + 1 \quad if j = k \\ a_k^* = a_k \quad else if j \neq k \end{cases}$$

$$b_k^* = \begin{cases} b_k^* = b_k + 1 \quad if j > k > 0 \\ b_k^* = b_k \quad else if j \leq k \end{cases}$$

Proof. We first note that according to our model $P(X = j|P) = p_j$.
Therefore

$$\begin{aligned} f(P|X = j) &= \frac{f(X = j|P) * f(P)}{\int_p f(X = j|P) * f(P) dP} && \text{Using the Bayes theorem} \\ &\propto f(X = j|P) * f(P) \\ &\propto p_j * \{\prod_{i=1}^{N-1} \mathcal{B}(a_i, b_i)] p_N^{b_{N-1}-1} \prod_{i=1}^{N-1} [p_i^{a_i-1} (\sum_{k=i}^{N} p_k)^{b_{i-1}-(a_i+b_i)}\} && \text{Using Theorem 1} \end{aligned}$$

Let $P_i = \sum_{k=i}^{N} p_k$. then we can say.

$$f(P|X = j) \propto \{p_N^{b_{N-1}-1} \prod_{i=1}^{j-1} p_i^{a_i-1} (P_i)^{(b_{i-1}+1)-(a_i+(b_i+1))}\} p_j^{(a_j+1)-1} P_j^{(b_{j-1}+1)-((a_j+1)+(b_j))}\}$$
$$\{\prod_{i=1}^{N-1} p_i^{a_i-1} (P_i)^{b_{i-1}-(a_i+b_i)}\} \tag{5.17}$$

From the above we can deduce that the vector P is distributed $\mathcal{G}(a_1^*, b_1^*, ..., a_N^*, b_N^*)$.

Which implies

$$p_1 = V_1^* \quad and \quad p_k = (1 - V_1^*)...(1 - V_{k-1}^*)V_k^* \tag{5.18}$$

where

$$(V_k^*|X_j) \sim \beta(a_k^*, b_k^*) \tag{5.19}$$

$$a_k^* = \begin{cases} a_k^* = a_k + 1 \quad if j = k \\ a_k^* = a_k \quad else if j \neq k \end{cases}$$

$$b_k^* = \begin{cases} b_k^* = b_k + 1 \quad if j > k > 0 \\ b_k^* = b_k \quad else if j \leq k \end{cases}$$

□

Proposition 23. *conditional for γ is*

$$f(\gamma|p) = \Gamma(N + \epsilon_1 - 1, \epsilon_2 - \sum_{i=1}^{N-1} \log(1 - V_i^*)) \tag{5.20}$$

Proof. We can observe from Theorem 1 that the distribution for P is

$$\{\prod_{i=1}^{N-1} \mathcal{B}(a_i, b_i)] p_N^{b_{N-1}-1} \prod_{i=1}^{N-1} [p_i^{a_i-1} (\sum_{k=i}^{N} p_k)^{b_{i-1}-(a_i+b_i)}\} \tag{5.21}$$

so now using the fact that we have $a_i = 1$and $b_i = \gamma$ and also using the fact that $\Gamma(1 + \gamma) = \gamma\Gamma(\gamma)$ we get that

$$f(p|\gamma) \propto \gamma^{N-1} p_N^{\gamma-1} = \gamma^{N-1} e^{(\gamma-1)\log(p_N)} \tag{5.22}$$

Therefore we can finally write using Bayes theorem

$$
\begin{aligned}
f(\gamma|p) &\propto f(p|\gamma)f(\gamma) \\
&\propto \gamma^{N-1}e^{(\gamma-1)\log(p_N)} * \{\gamma^{\epsilon_1-1}e^{-\epsilon_2\gamma}\} && \text{from the Gamma distribution} \\
&\propto \gamma^{N+\epsilon_1-2}e^{-(\epsilon_2-\log(p_N))\gamma}e^{-\log(p_N)} && \text{Its an un-normalized Gamma distribution} \\
f(\gamma|p) &= \Gamma(N+\epsilon_1-1, \epsilon_2-\log(p_N)) && \{p_N = \prod_{i=1}^{N-1} V_i^*\} \\
f(\gamma|p) &= \Gamma(N+\epsilon_1-1, \epsilon_2-\sum_{i=1}^{N-1}\log(1-V_i^*))
\end{aligned}
$$

□

5.3 Gibbs Sampler Algorithm

We give the whole algorithm as follows

Algorithm 2: Algorithm for Geometric levy process

Choose hyper parameters ϵ_1, ϵ_2 ;
Generate γ from Gamma(ϵ_1, ϵ_2)
Generate N paths of the Markov Chain.
Draw p=$p_1, p_2, ..., p_N$ from stick breaking mechanism(γ, H)
Choose one of the N paths according to the corresponding p
for $i = 1$ *to* 10000 **do**
    Update the conditional of X according to the conditional given in equation 5.14
    Choose one of the paths according to the distribution of X;
    Given a path X update the conditionals for all the parameters:
        Draw a new μ_j for all the states j according to the equation 5.10
        Draw a new δ_j for all the states j according to the equation 5.11
        Draw a new α according to the equation 5.12
        Draw a new β according to the equation 5.13
    Define a_k^* and b_k^* according to equation (5.16) where k is the index of chosen path
    Compute $p_i's$ according to the equation (5.15)
    Draw a new γ according to the equation (5.20)
end
Draw the most probable Markov Chain Path from the posterior distribution.
Estimate H according to the Maximum Likelihood Estimation of the most probable Markov Chain Path.

Chapter 6

Particle MCMC

As stated in the previous sections, the main challenge in inference for Markov Modulated Levy Processes is to sample effectively from the distribution of the Underlying Markov Chain. Therefore we used Truncated Dirichlet Prior method in the previous sections. In this section, We would like to introduce another method called the Particle Markov Chain Monte Carlo which solves the problem of sampling from the Underlying Markov Chain using Sequential Monte Carlo. The references used for this section are [Christophe Andrieu and Holenstein, 2010] and [Arnaud Doucet, 2011]. We first define the basic model which is very similar to the previous case but is defined again for clarity.

Definition 23. *We define $(X_t)_{t\in\{0,1,..,T\}}$ to be a Discrete time Markov chain on finite space $\mathcal{S}$: = 1, 2, ..., M. The probability transition matrix of Markov Chain, denoted by $\mathbb{P}$, is given by*

$$\mathbb{P}_{ij} \geq 0 \text{ for all } i,j \in \mathcal{S} \text{ and } \mathbb{P}_{ij} = P(X_{t+1} = j | X_t = i) \tag{6.1}$$

Next we consider some observable process say S observed at discrete times Δt for simplicity we assume $\Delta t = 1$. It has independent and stationary increments and has parameters θ which determine its distribution. Mathematically,

$$S_{t+1} - S_t \sim f(\theta) \tag{6.2}$$

Where $f(\theta)$ denotes the density of the increments. Now we want the parameters denoted by θ to be regime switching or more precisely each state of the above defined

continuous time Markov chain corresponds to a different set of parameters θ_k. If lets say the dimension of θ is K, i.e. the distribution f is a function of K variables, and our Markov Chain has say M states. Then we can denote it as

$$\Theta : \begin{bmatrix} 1 \\ \vdots \\ M \end{bmatrix} \longrightarrow \left\{ \begin{bmatrix} \theta_{11} \\ \vdots \\ \theta_{1K} \end{bmatrix}, ..., \begin{bmatrix} \theta_{M1} \\ \vdots \\ \theta_{MK} \end{bmatrix} \right\} \tag{6.3}$$

where

$$\Theta(X_t) \longrightarrow \begin{bmatrix} \theta_{X_t 1} \\ \vdots \\ \theta_{X_t K} \end{bmatrix} \tag{6.4}$$

So we can observe that θ_{ij} corresponds to value of the j^{th}parameter of the density in state i.
So now our model can be denoted in the mathematical form as:

$$S_{t+1} - S_t \sim f(\Theta(X_t))$$

If the Markov chain was in state X_t at time t .
Note that now the distribution of increments is not identical or stationary anymore but still independent.
Now, we want to fit a regime-switching model such as above. Thus the optimal set of parameters to estimate is $(\hat{\theta}_{i,j}\, \hat{X}) \in \mathbb{R}^{M\times K} \times \mathcal{D}$. where $i = 1, ..., M$ and $j = 1, ..., K$. Where $\{\hat{\theta}_{i,j}; i = 1, ..., M; j = 1, ..., K\} \in \mathbb{R}^{M\times K}$ are the parameters for the density associated and $\hat{X} \in \mathcal{D}$ is the path taken by the underlying Markov Chain and $\mathcal{D}$ is the set of paths of the discrete time Markov Chain
We define a prior density p_θ for θ and Bayesian inference relies on the joint density

$$f(\Theta; X_{\{1:T\}}|S_{\{1:T\}}) \propto f_\theta(X_{\{1:T\}}|S_{\{1:T\}})f(\theta)$$

For non-linear non-Gaussian models, $f_\theta(X_{\{1:T\}}|S_{\{1:T\}})$ and $f(\theta; X_{\{1:T\}}|S_{\{1:T\}})$ do not usually admit closed form expressions, making inference difficult in practice. It is therefore, necessary to resort to approximations. Monte Carlo methods have been

shown to provide a flexible framework to carry out inference in such models.

6.1 Particle Gibbs Sampler

Our aim is to find the posterior $f(\Theta; X_{\{1:T\}}|S_{\{1:T\}})$, We will be using a modified Gibbs sampler to do that. This works by alternately sampling from the conditional distributions to get the joint distribution. So, we need to sample from:

$$f_\theta(X_{\{1:T\}}|S_{\{1:T\}})$$
$$f(\theta|X_{\{1:T\}}, S_{\{1:T\}})$$

We use technique called Sequential Monte Carlo to sample from $f_\theta(X_{\{1:T\}}|S_{\{1:T\}})$

6.1.1 Sequential Monte Carlo

Sequential Monte Carlo are used to approximate posterior densities $\{f_\theta(X_{\{1:n\}}|S_{\{1:n\}}); n \geq 1\}$ sequentially. It provides an approximation of the posterior density as a set of $N \geq 1$ weighted random particles which can be represented as :

$$\hat{f}_{\theta,S_{1:T}}(dx_{1:T}) := \sum_{k=1}^{N} W_T^k \delta_{X_{1:T}^k}(dx_{1:T})$$

Where each $X_{1:n}^k$ particle is a sample Markov Chain path from time 1 to n. And W_n^k is the weight associated with the k^{th} particle at time n. The particles $X_{1:n}^k$ and the weights W_n^k are random and their distribution depend on θ and observations $S_{1:T}$. Sequential Monte Carlo (SMC) relies on a proposal density function which is chosen by the user, denoted by $h_\theta(x_{1:n}|S_{1:n})$ which has the following form

$$h_\theta(x_{1:T}|S_{1:T}) = h_\theta(x_1|S_1) \prod_{n=2}^{T} h_\theta(x_n|S_n, x_{n-1})$$

This function captures the probability of a sequence of states (x_1 to x_n) given a sequence of observations (S_1 to S_n). The important thing to account for while specifying proposal density is that the support of the target or posterior density should be a subset of support of the proposal density. Thus usually proposal densities are chosen with a wide support.

The key point is that the proposal density allows us to sample states sequentially. It does this by considering two parts:

- At time n=1, we sample the initial state x_1 from a function $h_\theta(x_1|S_1)$that depends only on the first observation S_1.

- For time n=2,...,T we sample the next state x_n using a function $h_\theta(x_n|S_n, x_{n-1})$ that considers both the current observation S_n and the previous state X_{n-1}.

We also specify the prior for the Markov chain. Which for our use will be as specified as in previous sections as:

- The initial distribution $\pi_1(.) = (1/M, ..., 1/M)$.

- The transition matrix according to the prior distribution is an uniform matrix as $\mathbb{P}$ with $\mathbb{P}_{ij} = 1/M$ for all $i, j = 1...M$

We then use importance sampling method where we find the ratio of true density and proposal density to approximate the distribution of particles at each step which are called the weights associated with each particle.
We denote the weight associated with particle k at time 1 i.e. X_1^k with W_1^k. Therefore the weights can be specified as :
for t=1

$$W_1^k = \frac{f_\theta(X_1^k|S_1)}{h_\theta(X_1^k|S_1)} = \frac{\pi_1(X_1^k)f_\theta(S_1|X_1^k)}{f_\theta(S_1)h_\theta(X_1^k|S_1)} = \frac{(1/M)f_\theta(S_1|X_1^k)}{f_\theta(S_1)h_\theta(X_1^k|S_1)}$$

since $f_\theta(S_1)$ doesn't depend on X_1 we can say that

$$W_1^k \propto \frac{(1/M)f_\theta(S_1|X_1^k)}{h_\theta(X_1^k|S_1)}$$

or the un-normalized weights can be calculated as :

$$\tilde{w}_1^k = \frac{(1/M)f_\theta(S_1|X_1^k)}{h_\theta(X_1^k|S_1)} \tag{6.5}$$

Similarly at time n=2,...,T , we can calculate the un-normalized weight associated with each particle as

$$\begin{aligned}
\tilde{w}_n^k &= \frac{f_\theta(X_{1:n}^k, S_{1:n})}{f_\theta(X_{1:n-1}^{A_{n-1}^k}, S_{1:n-1}) h_\theta(X_n^k | S_n, X_{n-1}^{A_{n-1}^k})} \\
&= \frac{f_\theta(X_{1:n}^k, S_{1:n} | X_{1:n-1}^{A_{n-1}^k}, S_{1:n-1})}{h_\theta(X_n^k | S_n, X_{n-1}^{A_{n-1}^k})} \\
&= \frac{f_\theta(X_n^k, S_n | X_{1:n-1}^{A_{n-1}^k}, S_{1:n-1})}{h_\theta(X_n^k | S_n, X_{n-1}^{A_{n-1}^k})} \\
&= \frac{f_\theta(X_n^k, S_n | X_{1:n-1}^{A_{n-1}^k})}{h_\theta(X_n^k | S_n, X_{n-1}^{A_{n-1}^k})} \quad S_n \text{ depends only on } X_n \\
&= \frac{f_\theta(X_n^k | X_{n-1}^{A_{n-1}^k}) \times f_\theta(S_n | X_n^k)}{h_\theta(X_n^k | S_{1:n}, X_{n-1}^{A_{n-1}^k})} \quad \text{using Markov property} \\
&= \frac{\mathbb{P}_{A_{n-1}^k, k} \times f_\theta(S_n | X_n^k)}{h_\theta(X_n^k | S_{1:n}, X_{n-1}^{A_{n-1}^k})} \quad \text{using prior defined previously} \qquad (6.6)
\end{aligned}$$

Where $\tilde{w}_n^k$ is the un-normalized weight associated with article k at time n i.e. the particle $X_{1:n}^k$ and therefore W_n^k is the weights associated with the corresponding particle. A_{n-1}^k is used to denote which particle (identified by its index) was the "parent" of particle X_n^k at the previous time step (n-1). We then normalize the above computed weights to obtain W_n^k the final weights. We also specify the following notation which is important in understanding the working of the subsequent defined methods. B_n^k denotes the index of the ancestor of particle k at generation n. So we can observe that $B_T^k = k$. Following the above notation we can write the evolution of a particle from time 1 to T as follows: $X_{1:T}^k = (X_1^{B_1^k}, X_2^{B_2^k}, ..., X_{T-1}^{B_{T-1}^k}, X_T^{B_T^k})$ and we have the following ancestral lineage associated with the particle chain $B_{1:T}^k = (B_1^k, ..., B_{T-1}^k, B_T^k = k)$. Here's how the sequential Monte Carlo algorithm works:

- We first sample N particles for time 1 using the proposal density $h_\theta(.|S_1)$.
- we then find the un-normalized weights associate with each particle as specified in (6.5) and normalize them.
- We then specify $\hat{f}_1(dx_1) := \sum_{k=1}^N W_1^k \delta_{X_1^k}(dx_1)$ which is the distribution of particles as time 1 and also $\hat{\mathcal{F}}_1(.) := \sum_{k=1}^N W_1^k \delta_k(.)$ which is the distribution over

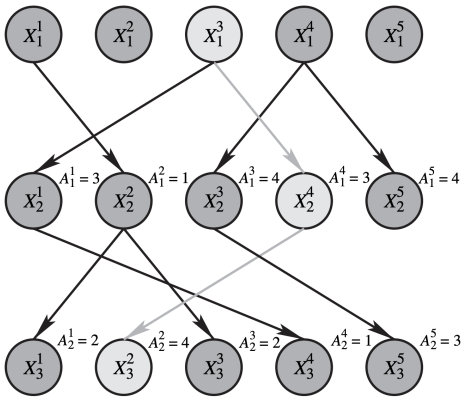


Fig. 6.1: Example of SMC algorithm with N=5 and T=3, the lighter path is $X_{1:3}^2 = (X_1^3, X_2^4, X_3^2)$ and has the lineage $B_{1:3}^2 = (3, 4, 2)$, taken from [Christophe Andrieu and Holenstein, 2010]

the indexes of the particles at time 1.

- We then follow the following loop for n=2 to T:

  - We choose parent indexed A_{n-1}^k according to the distribution defined at time n-1 i.e. $\hat{\mathcal{F}}_{n-1}(.) := \sum_{k=1}^N W_{n-1}^k \delta_k(.)$.

  - Then sample N particles for time n using the proposal density $h_\theta(.|S_n, X_{n-1}^{A_{n-1}^k})$ and define the k^{th} particle at time n to be $X_{1:n}^k := (X_{1:n-1}^{A_{n-1}^k}, X_n^k)$

  - We then find the un-normalized weights associate with each particle as specified in (6.6) and normalize them.

  - We then specify $\hat{f}_n(dx_{1:n}) := \sum_{k=1}^N W_n^k \delta_{X_{1:n}^k}(dx_{1:n})$ which is the distribution of particles as time n and also $\hat{\mathcal{F}}_n(.) := \sum_{k=1}^N W_n^k \delta_k(.)$ which is the distribution over the indexes of the particles at time n.

We can then approximate the conditional distribution of

$$f_\theta(X_{\{1:T\}}|S_{\{1:T\}}) \sim \hat{f}_T(dx_{1:T}) := \sum_{k=1}^N W_T^k \delta_{X_{1:T}^k}(dx_{1:T})$$

The sequential monte carlo can be given as follows in algorithim form:

Algorithm 3: Sequential Monte Carlo

Sample X_1^k from the proposal density $h_\theta(.|S_1)$ for all $k \in \{1, ..., N\}$

Compute the weights associated with each particle as:

$$\tilde{w}_1(X_1^k) = \frac{f_\theta(X_1^k, S_1)}{h_\theta(X_1^k|S_1)} = \frac{\pi(X_1^k) f_\theta(S_1|X_1^k)}{h_\theta(X_1^k|S_1)}$$

Normalize the weights

$$W_1(X_1^k) := \frac{\tilde{w}_1(X_1^k)}{\sum_{l=1}^N \tilde{w}_1(X_1^l)}$$

Set the distribution as follows:

$$\hat{f}_1(dx_1) := \sum_{k=1}^N W_1^k \delta_{X_1^k}(dx_1)$$

$$\hat{\mathcal{F}}_1(.) := \sum_{k=1}^N W_1^k \delta_k(.)$$

for $n = 2$ *to* T **do**

  for $k = 1$ *to* N **do**

    Sample A_{n-1}^k from $\hat{\mathcal{F}}_{n-1}(.|\mathbf{W_{n-1}})$

    Sample X_n^k from $h_\theta(.|S_n, X_{n-1}^{A_{n-1}^k})$ and define $X_{1:n}^k := (X_{1:n-1}^{A_{n-1}^k}, X_n^k)$.

  end

  Compute the weight associated with each particle:

$$\tilde{w}_n(X_{1:n}^k) = \frac{f_\theta(X_{1:n}^k, S_{1:n})}{f_\theta(X_{1:n-1}^{A_{n-1}^k}, S_{1:n-1}) h_\theta(X_n^k|S_n, X_{n-1}^{A_{n-1}^k})} = \frac{f_\theta(X_n^k|X_{n-1}^{A_{n-1}^k}) \times f_\theta(S_n|X_n^k)}{h_\theta(X_n^k|S_{1:n}, X_{n-1}^{A_{n-1}^k})}$$

  Normalize the weights

$$W_n^k := \frac{\tilde{w}_n(X_{1:n}^k)}{\sum_{l=1}^N \tilde{w}_n(X_{1:n}^l)}$$

  Set the distribution as follows:

$$\hat{f}_n(dx_{1:n}) := \sum_{k=1}^N W_n^k \delta_{X_{1:n}^k}(dx_{1:n})$$

$$\hat{\mathcal{F}}_n(.) := \sum_{k=1}^N W_n^k \delta_k(.)$$

6.1.2 Conditional Sequential Monte Carlo

To use Sequential Monte Carlo as part of Gibbs sampler we need a modified version called as Conditional Sequential Monte Carlo. CSMC works by keeping a specific path through the states, denoted by $X_{1:T}$, and its associated lineage, denoted by$B_{1:T}$, fixed during the resampling step.
In contrast to the fixed path and lineage, all other particles (representing different potential paths through the data) are resampled as usual in Sequential Monte Carlo (SMC). This selective resampling allows CSMC to focus on refining the neighborhood around the chosen path, potentially leading to more efficient sampling in problems with regime switching models.

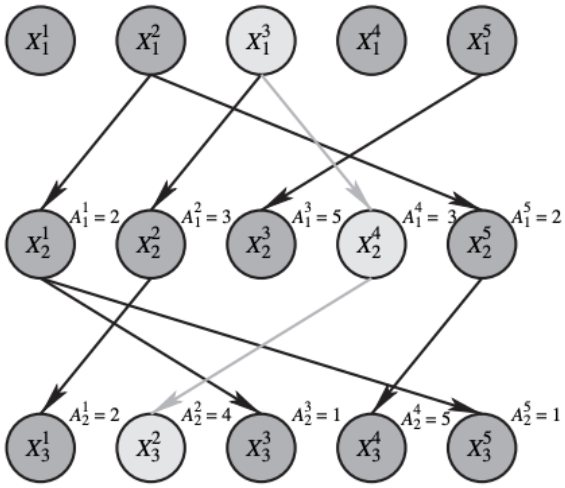


Fig. 6.2: Example of N=5 ancestral lineages generated by a conditional SMC algorithm for N =5 and T =3 conditional on $X_{1:3}^2$and $B_{1:3}^2$, taken from [Christophe Andrieu and Holenstein, 2010]

We give the algorithm for conditional sequnetial monte carlo as follows with number of particles N and time steps as 1 to T:

Algorithm 4: Conditional Sequential Monte Carlo

Define a specified path $X_{1:T} = (X_1^{B_1}, ..., X_T^{B_T})$ and the ancestral lineage $B_{1:T}$.

if $k \neq B_1$ **then**

  Sample X_1^k *from the proposal density* $h_\theta(.|S_1)$ *for all* $k \in \{1, ..., N\}$

end

Compute the weights associated with each particle as:

$$\tilde{w}_1(X_1^k) = \frac{f_\theta(X_1^k, S_1)}{h_\theta(X_1^k|S_1)} = \frac{\pi(X_1^k) f_\theta(S_1|X_1^k)}{h_\theta(X_1^k|S_1)}$$

Normalize the weights

$$W_1(X_1^k) := \frac{\tilde{w}_1(X_1^k)}{\sum_{l=1}^N \tilde{w}_1(X_1^l)}$$

Set the distribution as follows:

$$\hat{f}_1(dx_1) := \sum_{k=1}^N W_1^k \delta_{X_1^k}(dx_1), \quad \hat{\mathcal{F}}_1(.) := \sum_{k=1}^N W_1^k \delta_k(.)$$

for $n = 2$ *to* T **do**

  for $k = 1$ *to* N **do**

    if $k \neq B_n$ **then**

      Sample A_{n-1}^k *from* $\hat{\mathcal{F}}_{n-1}(.|\mathbf{W_{n-1}})$

      Sample X_n^k *from* $h_\theta(.|S_n, X_{n-1}^{A_{n-1}^k})$ *and define* $X_{1:n}^k := (X_{1:n-1}^{A_{n-1}^k}, X_n^k)$.

    end

  end

  Compute the weight associated with each particle:

$$\tilde{w}_n(X_{1:n}^k) = \frac{f_\theta(X_{1:n}^k, S_{1:n})}{f_\theta(X_{1:n-1}^{A_{n-1}^k}, S_{1:n-1}) h_\theta(X_n^k|S_n, X_{n-1}^{A_{n-1}^k})} = \frac{f_\theta(X_n^k|X_{n-1}^{A_{n-1}^k}) \times f_\theta(Sn|X_n^k)}{h_\theta(X_n^k|S_{1:n}, X_{n-1}^{A_{n-1}^k})}$$

  Normalize the weights

$$W_n^k := \frac{\tilde{w}_n(X_{1:n}^k)}{\sum_{l=1}^N \tilde{w}_n(X_{1:n}^l)}$$

  Set the distribution as follows:

$$\hat{f}_n(dx_{1:n}) := \sum_{k=1}^N W_n^k \delta_{X_{1:n}^k}(dx_{1:n}), \quad \hat{\mathcal{F}}_n(.) := \sum_{k=1}^N W_n^k \delta_k(.)$$

6.1.3 Particle Gibbs

Since the conditionals for the invidual parameters of $\Theta = \{\mu_i, \delta_i, \alpha, \beta \ \ ; i = 1, ..., M\}$ for MMLPS and $\Theta = \{\kappa_i, \theta_i, \sigma_i \ \ ; i = 1, ..., M\}$ for MMOU, will be same as that in previous sections(Chapter 4 for MMOU and chapter 5 for MMLP). this section focuses on the generalized Particle Gibbs Sampler (PG) algorithm applicable to a broader class of regime-switching models.
Here is the Particle Gibbs sampler for a general regime switching model.

Algorithm 5: Particle Gibbs Sampler

1 Place appropriate prior for each component of Θ
2 Sample $\theta(0)$ from the priors.
3 Generate one sample path of the Markov Chain from the uniform prior, say $X_{1:T}(0)$.
4 **for** $i = 1$ *to* . **do**
5 Generate N paths according to conditional sequential Monte Carlo conditioned on $\theta(i-1)$ and $X_{1:T}(i-1)$. Set the final distribution obtained as $\{\hat{f}_T^i(dx_{1:T}) := \sum_{k=1}^N W_T^k \delta_{X_{1:T}^k}(dx_{1:n})\}$
6 Choose one of the paths according to the distribution of $X_{1:T}(i)$ from the distribution obtained above i.e $\{\hat{f}_T^i(dx_{1:T})\}$.
7 **Given a path** $X_{1:T}(i)$
8 Sample from the conditionals of each component of θ
9 **end**

The algorithm in the following manner:
It starts by defining priors for all parameters in the model denoted by θ. We also initialise values of parameters using samples from the prior distribution.
Then, it generates an initial sample path for the hidden Markov chain $X_{1:T}(0)$ using the uniform prior defined previously. We then iterate over a loop and within each iteration:

- We first generate N new sample paths and distribution $\{\hat{f}_T^i(dx_{1:T})\}$ over these paths using Conditional SMC conditioned on the current state of the model parameters, $\theta(i-1)$ and the previous hidden state path,$X_{1:T}(i-1)$.

- One path $X_{1:T}(i)$ is chosen from the generated set based on the distribution $\{\hat{f}_T^i(dx_{1:T})\}$ over the paths as defined by the CSMC.

- Gibbs Sampling: Given the chosen hidden path, $X_{1:T}(i)$ we use Gibbs sampling by sampling from conditional distributions of individual components of the model parameters θ.

Using the CSMC sampler we are able to draw from an approximation of $f_\theta(X_{1:T}|S_{1:T})$ and then we use Gibbs sampler to draw samples from the distribution $f(\theta|X_{1:T}, S_{1:T})$. By iteratively performing these steps, the PG sampler explores the joint posterior distribution of all model parameters and the hidden states and the algorithm should converge to the joint distribution $f(\theta, X_{1:T}, S_{1:T})$.

Chapter 7

Results

In this section, we present the results of the algorithm on simulated data for both the kind of models discussed in the previous section.

7.1 Markov Modulated Ornstein-Uhlbeck Process

We simulated a Markov Modulated OU process where all the three parameters κ, σ, θ were dependent on underlying Markov Chain. We took the underlying Markov Chain to have two states $\{0, 1\}$ with the transition matrix as:

$$\begin{bmatrix} 0.55 & 0.45 \\ 0.35 & 0.65 \end{bmatrix}$$

Also the values for the parameters associated with each state is as follows:

$$\kappa = [0.1, 0.5]$$
$$\theta = [1, 2]$$
$$\sigma = [1.2, 1.5]$$

The resulting Markov Chain and also the resulting Markov Modulated Ornstein Uhlbeck process is as follows:

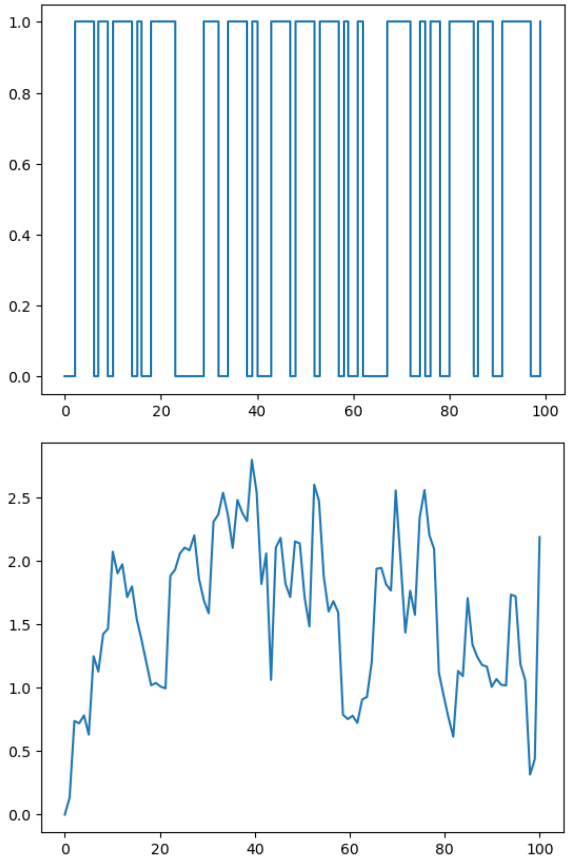
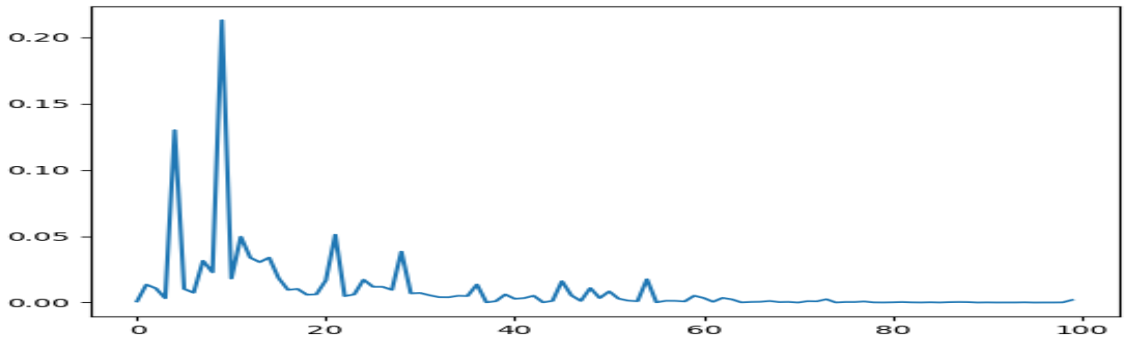


Fig. 7.1: a)Markov Chain with transition Matrix=[[0.55 ,0.45],[0.35 , 0.65] b)the resulting OU process

7.1.1 Truncated Dirichlet Prior Method

We generated 100 Markov Chain paths with uniform Transition Matrix and uniform initial distribution as stated above. Then we placed a stick breaking Truncated Dirichlet prior with number of classes equal to 100. Which gives the initial distribution as follows over the 100 MC paths:



Then we defined the priors for each parameter as described in the sections above. We started with an initial value of 1 for all the parameters and ran our algorithm 2000 times. We took a burn-in period of 1000 iterations. The algorithm estimate the parameters as follows:

	κ_0	κ_1	θ_0	θ_1	σ_0	σ_1
True value	0.1	0.5	1	2	1.2	1.5
Posterior Mode	0.09	0.41	1.1	2	1.2	1.5

Table 7.1: True and estimated values

The posterior distribution and the trace plots for each of the above parameters is as follows:

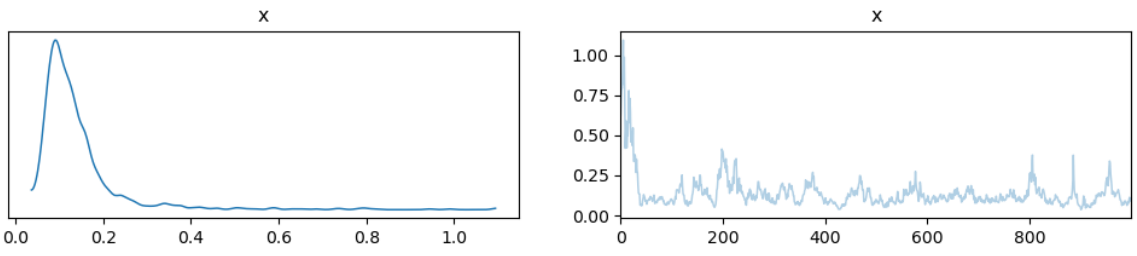


Fig. 7.2: Kappa in state 0

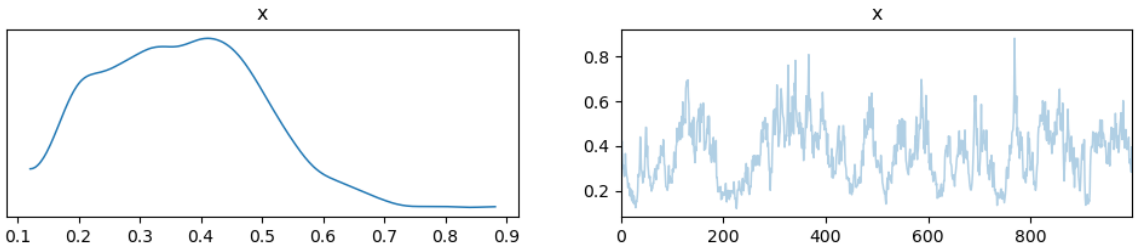


Fig. 7.3: Kappa in state 1

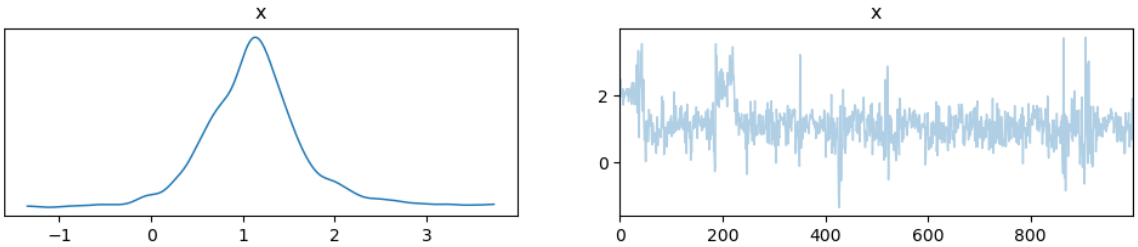


Fig. 7.4: Theta in state 0

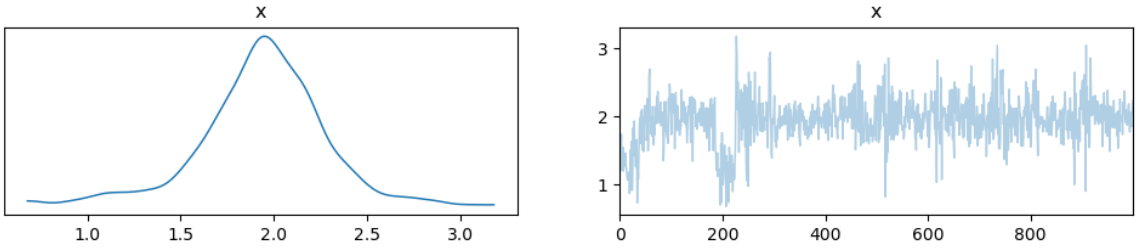


Fig. 7.5: Theta in state 1

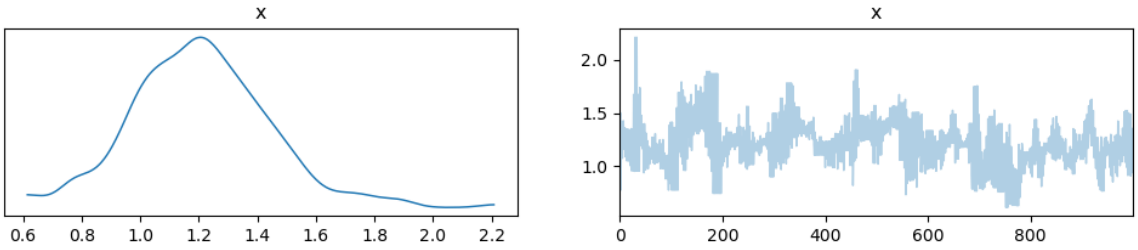


Fig. 7.6: Sigma in state 0

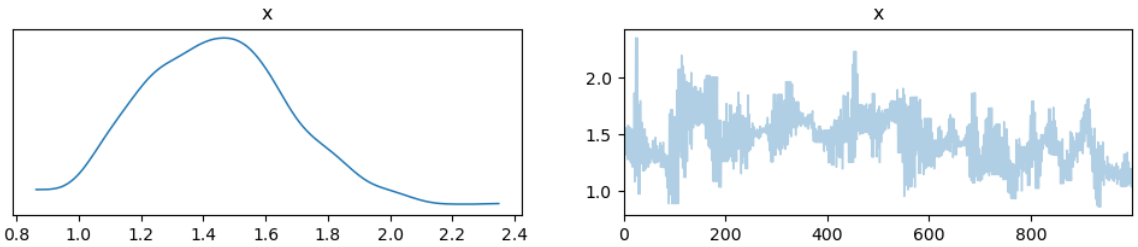


Fig. 7.7: Sigma in state 1

Also we estimate the transition matrix from the most probable path as follows:

$$\begin{bmatrix} 0.667 & 0.334 \\ 0.314 & 0.686 \end{bmatrix}$$

7.1.2 Particle MCMC method

We then used the same data to also try our Particle Gibbs method on the Markov Modulated OU process. The proposal densities $h_\theta(x_n|S_n, x_{n-1})$ were taken to be proportional to the likelihood function i.e $h_\theta(x_n|S_n, x_{n-1}) \propto f_\theta(S_n|x_n)$. which we then normalize.

Then we defined the priors for each parameter as described in the sections above. We started with an initial value of 1 for all the parameters and initial Markov chain was generated from the uniform prior over path space. We ran our algorithm 1000 times. We took a burn-in period of 900 iterations. The algorithm estimate the parameters as follows:

	κ_0	κ_1	θ_0	θ_1	σ_0	σ_1
True value	0.1	0.5	1	2	1.2	1.5
Posterior Mode	0.12	0.55	1.3	2.2	1.1	1.4

Table 7.2: True and estimated values

The posterior distribution and the trace plots for each of the above parameters is as follows:

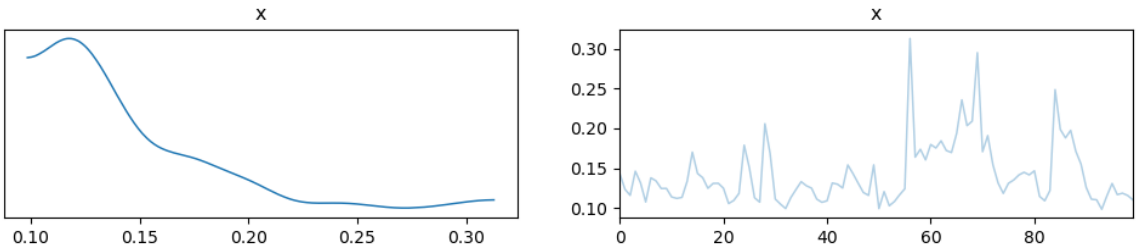


Fig. 7.8: Kappa in state 0

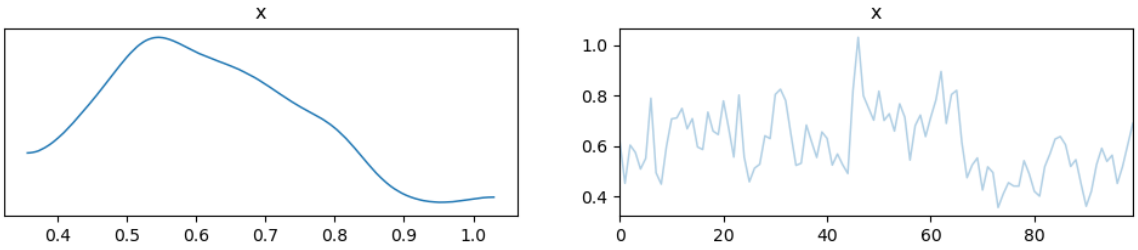


Fig. 7.9: Kappa in state 1

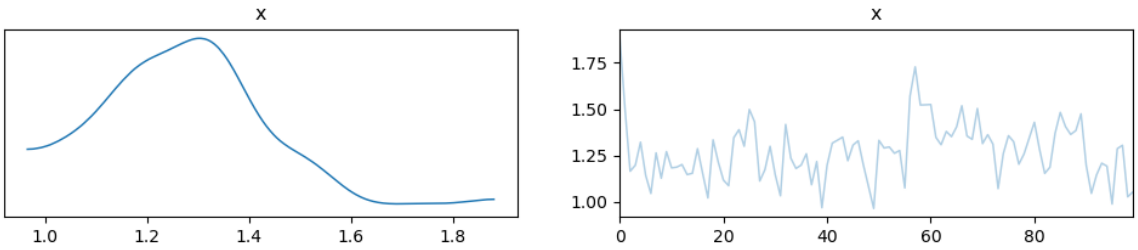


Fig. 7.10: Theta in state 0

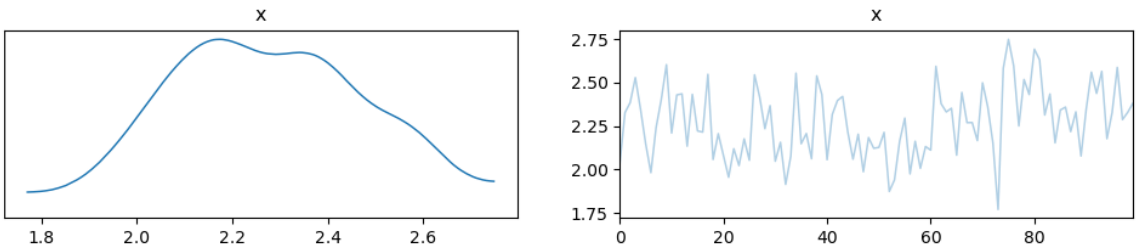


Fig. 7.11: Theta in state 1

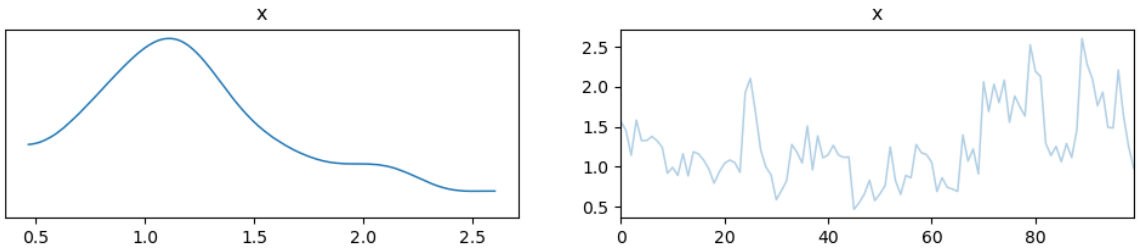


Fig. 7.12: Sigma in state 0

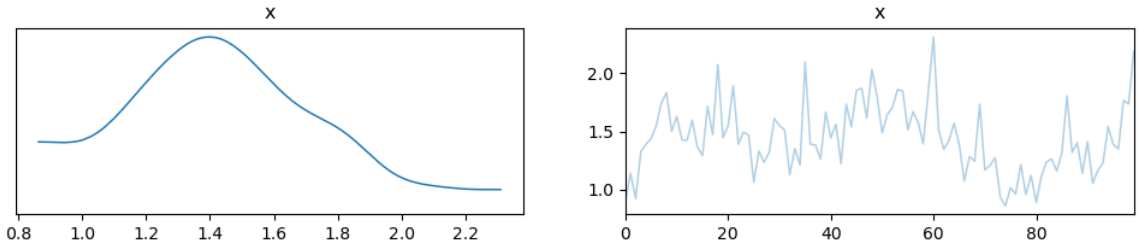


Fig. 7.13: Sigma in state 1

7.2 Markov Modulated NIG Levy process

7.2.1 Truncated Dirichlet Prior Method

We simulated a Markov Modulated NIG Levy process where all the parameters μ, δ were dependent on underlying Markov Chain. We took the underlying Markov Chain to have two states $\{0, 1\}$ with the transition matrix as:

$$\begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$$

Also the values for the parameters associated with each state is as follows:

$$\mu = [1, 1.3]$$
$$\delta = [1.2, 1.7]$$

The resulting Markov Chain and also the resulting Markov Modulated Levy process is as follows:

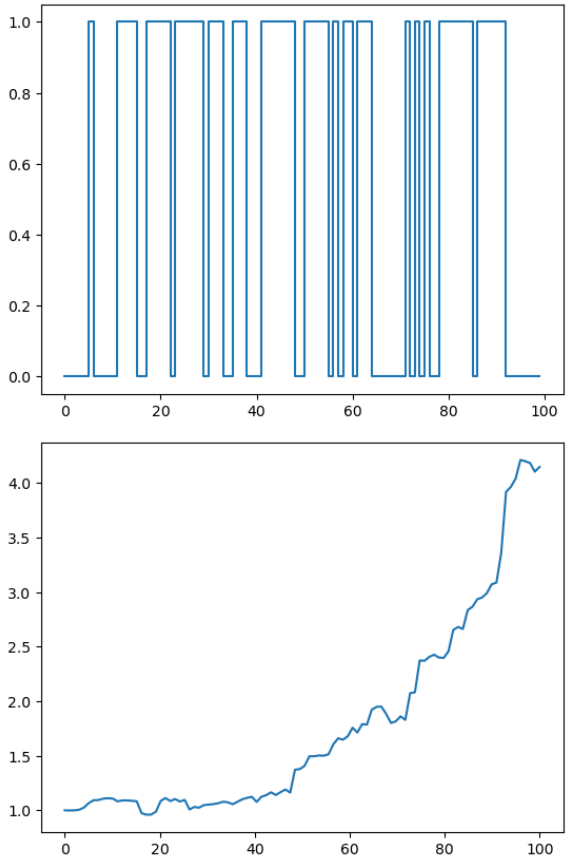


Fig. 7.14: a)Markov Chain with transition Matrix=[[0.6, 0.4], [0.3, 0.7]] b)the resulting Levy process

Then we defined the priors for each parameter as described in the sections above. We started with an initial value of 1 for δ for both states and initial value of 0 for μ for both states ,ran our algorithm 2000 times. We took a burn-in period of 1000 iterations. The algorithm estimate the parameters as follows:

	μ_0	μ_1	δ_0	δ_1
True value	1	1.3	1.2	1.7
Posterior Mode	1	1.3	1.2	1.65

Table 7.3: True and estimated values

The posterior distribution and the trace plots for each of the above parameters is as follows:

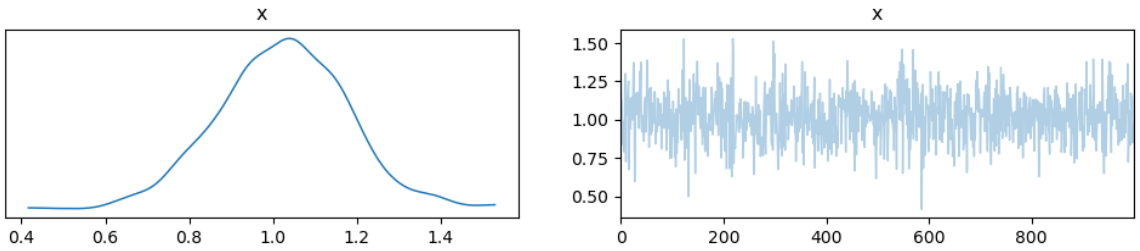


Fig. 7.15: μ in state 0

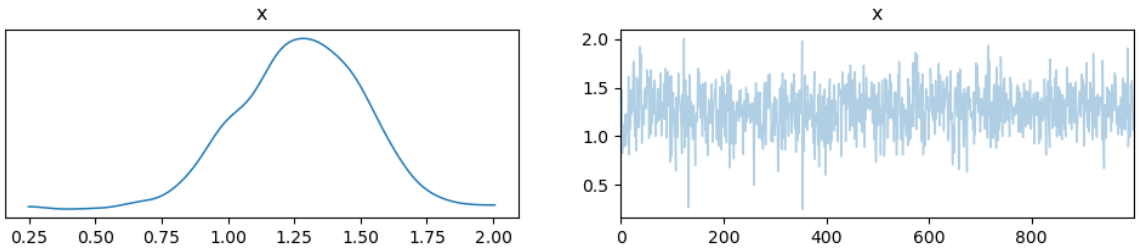


Fig. 7.16: μ in state 1

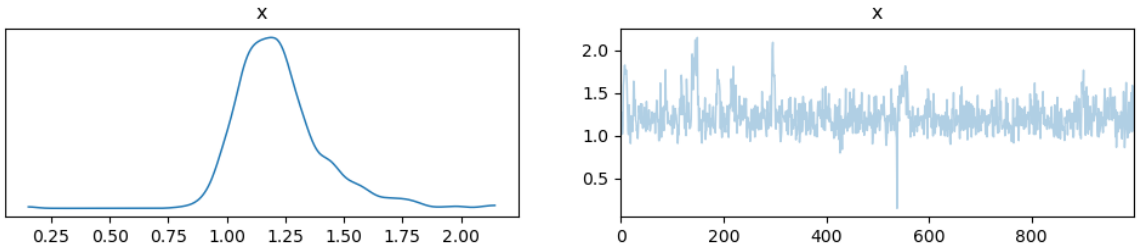


Fig. 7.17: δ in state 0

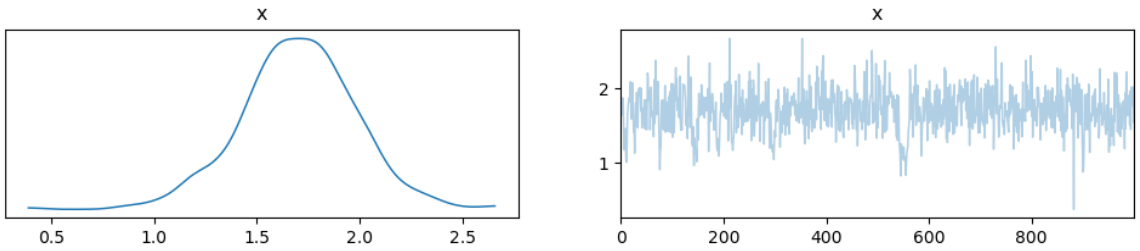


Fig. 7.18: δ in state 1

Also we estimate the transition matrix from the most probable path as follows:

$$\begin{bmatrix} 0.596 & 0.404 \\ 0.281 & 0.719 \end{bmatrix}$$

7.2.2 Particle MCMC method

We then used the same data to also try our Particle Gibbs method on the Markov Modulated NIG-Levy process. The proposal densities $h_\theta(x_n|S_n, x_{n-1})$ were taken to be proportional to the likelihood function as before i.e $h_\theta(x_n|S_n, x_{n-1}) \propto f_\theta(S_n|x_n)$. which we then normalize.
Then we defined the priors for each parameter as described in the sections above. We started with an initial value of 1 for δ for both states, initial value of 0 for μ for both states and initial Markov chain path was generated from the uniform prior over path space. We ran our algorithm 1000 times. The algorithm estimate the parameters as follows:

	μ_0	μ_1	δ_0	δ_1
True value	1	1.3	1.2	1.7
Posterior Mode	1.1	1.4	1.19	1.5

Table 7.4: True and estimated values

The posterior distribution and the trace plots for each of the above parameters is as follows:

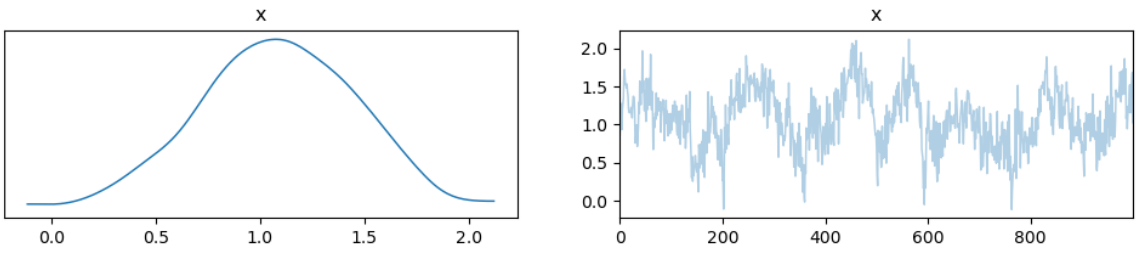


Fig. 7.19: μ in state 0

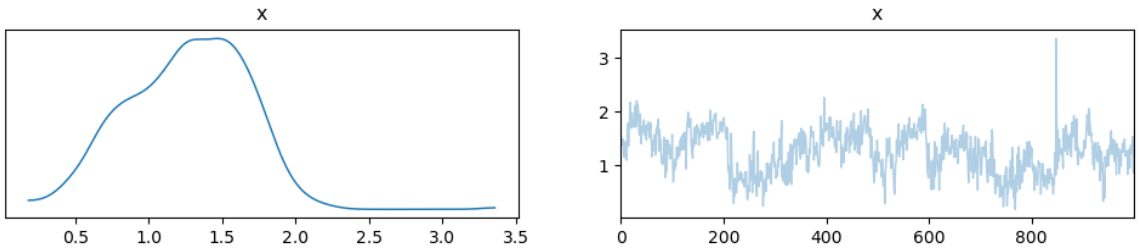


Fig. 7.20: μ in state 1

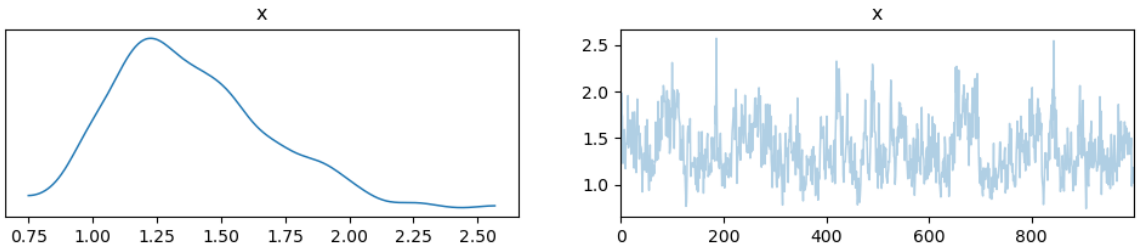


Fig. 7.21: δ in state 0

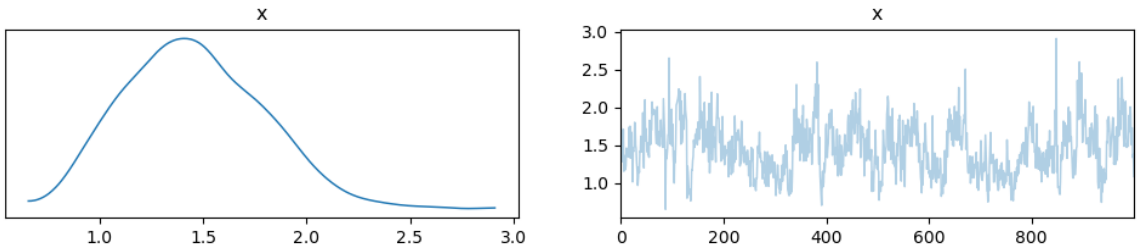


Fig. 7.22: δ in state 1

Chapter 8

Conclusion

The interplay between Markov Modulated Levy processes offers a powerful framework for modeling real-world phenomena exhibiting dynamic jumps and state-dependent behavior. Although a lot of literature exists on using MMLPs for derivative pricing and risk management. Only one approximate Expectation Maximization approach exists for performing Inference for Markov Modulated Levy Process.
In this project we set to develop an efficienct way to perform Inference for MMLPs when both the parameters and the underlying Markov chain is unknown.
We used Bayesian inference as it provides mathematical rigorous approach and has the following advantages over the Frequentest approach: **Uncertainty Quantification**: It provides us with a posterior distribution which can be used to construct confidence intervals allowing us to perform uncertanity quantification as well unlike point estimates.
Incorporating Prior Knowledge: Using the prior distribution allows us to incorporate prior information in our inference procedure like the possible ranges of estimators and helps to perform robust inference even in presence of outliers.

To carry out Bayesian inference we needed to find the conditional distribution of each parameter for our model. The main challenge to that was we considered Markov Chain paths as parameters and the space of Markov Chain paths is a complex parameter space and sampling from such a space is difficult. Therefore in order to tackle that challenge we used two different methods Truncated Dirichlet prior method and Particle MCMC. Which allowed us to sample effectively from the conditional distribution for Markov Chain paths.

We developed these new way of performing Bayesian inference for Markov Modulated Levy processes which can efficiently and precisely perform the inference giving us posterior distribution over both the parameter space and the Markov Chain path space.

To check for the effectiveness of the method we applied it to two models. First, a Markov modulated OU process. We considered the three parameters(κ, σ, θ) associated with OU process to switch between two states. Then we used our methods to estimate the parameters in each state. As we can see from the previous chapter the mode of posterior distribution is very close to the true values as well the posterior distribution provides provides confidence intervals as well. We also estimate the transition matrix associated with the underlying Markov Chain which also seems quite close to the true values.

Similarly we also simulated Markov Modulated exponential NIG process which switches between two states and the parameters (μ, δ) change according to the switch, and applied our methods to the observed data. We again find the methods can effectively estimate the parameter values for each state. We also calculated the transition matrix for the underlying Markov chain which also comes quite close to the true values.

From the above and previous sections we can observe that the methods is effective in estimating both the parameters as well as transition matrix for the underlying Markov chain and can even be extended to other Markov switching stochastic process. Although theses methods shows great promise in these simulation studies. We also need to check the effectiveness of the algorithm by applying to real financial data. Also the posterior distribution and therefore the confidence intervals calculated have a big spread. Since these experiments were carried out on machine with limited computing power. It can assumed that if the algorithm can be executed on a machine with much more computing power, so that the loops can be executed more number of times we will have a posterior distribution with much less variance.

This study looked into the usefulness of novel methods for parameter estimation in hidden Markov-switching models. We applied the strategy to two simulated processes: Markov-modulated Ornstein-Uhlenbeck (MMOU) and Markov-modulated exponential Normal Inverse Gaussian. The results seems promising and the approach was successful in recovering the true parameter values for each state in both processes. Furthermore, the computed transition matrices closely approximated the actual underlying Markov chains. Thus, these methods should be investigated with more em-

pirical data and computing power and so that we can have a good understanding of the effectiveness and applicability to real world. And also compare it to other methods which can be applied to similar problems.

Bibliography

[Anindya Goswami, 2019] Anindya Goswami, Omkar Manjarekar, A. R. (2019). Option pricing in a regime switching jump diffusion model. *arXiv preprint arXiv:1811.11379*.

[Arnaud Doucet, 2011] Arnaud Doucet, A. M. J. (2011). *A tutorial on particle filtering and smoothing: Fifteen years later*. Oxford, UK: Oxford University Press.

[Barndorff-Nielsen, 1997] Barndorff-Nielsen, O. (1997). Processes of normal inverse gaussian type. *Finance Stochast*, 2:41–68.

[Casarin, 2004] Casarin, R. (2004). Bayesian inference for generalised markov switching stochastic volatility models. *SSRN Electronic Journal*.

[Cheng-Der Fuh, 2012] Cheng-Der Fuh, Kwok Wah, R. H. I. H. R.-H. W. (2012). Option pricing with markov switching. *Journal of Data Science*, 10.

[Chevallier and Goutte, 2017] Chevallier, J. and Goutte, S. (2017). On the estimation of regime-switching lévy models. *Stud. Nonlinear Dyn. E.*, 21(1):3–29.

[Christophe Andrieu and Holenstein, 2010] Christophe Andrieu, A. D. and Holenstein, R. (2010). Particle markov chain monte carlo methods. *J. R. Statist. Soc*, 72(3):269–342.

[Connor and Mosimann, 1969] Connor, R. J. and Mosimann, J. E. (1969). Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, 64(325):194–206.

[Cont and Tankov, 2004] Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Process*. Chapman and Hall/CRC Financial Mathematical Series.

[Das and Goswami, 2019] Das, M. K. and Goswami, A. (2019). Testing of binary regime switching models using squeeze duration analysis. *Int. J. Financ. Eng.*, 6(1).

[Haugh, 2021] Haugh, M. (2021). A tutorial on markov chain monte-carlo and bayesian modeling. `https://ssrn.com/abstract=3759243`.

[Inamura, 2006] Inamura, Y. (2006). Estimating continuous time transition matrices from discretely observed data. *Bank of Japan Working Paper Series*.

[Ishwaran and James, 2001] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 64(453):96.

[Ishwaran and Zarepour, 2000] Ishwaran, H. and Zarepour, M. (2000). Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87.

[Ivanov, 2022] Ivanov, R. (2022). The risk measurement under the variance-gamma process with drift switching. *J. Risk Financial Manag.*

[Madan and Seneta, 1990] Madan, D. and Seneta, E. (1990). The variance gamma (v.g.) model for share market returns. *The Journal of Business, 63(4):511–524. http://www.jstor.org/stable/2353303*.

[Milan Kumar Das, 2023] Milan Kumar Das, A Goswami, S. R. (2023). Inference of binary regime models with jump discontinuities. *Sankhya B*, 85(1).

[Rasmus, 2016] Rasmus, S. (2016). *Derivative Prices for Models using Levy Processes and Markov Switching*. Phd thesis, Lund University, Example City, CA. Available at https://lup.lub.lu.se/search/publication/fd8801e9-6d2d-4c63-af27-f9d0f1b7c06f.

[Schoutens, 2003] Schoutens, W. (2003). *Levy Processes in Finance*. Wiley series in probability and statistics.

[Sethuraman., 1994] Sethuraman., J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650.

[Shalizi, 2009] Shalizi, C. (2009). Maximum likelihood estimation for markov chains. `https://www.stat.cmu.edu/~cshalizi/462/lectures/06/markov-mle.pdf`.

[Shaw, 2019] Shaw, C. (2019). Regime-switching and levy jump dynamics in option-adjusted spreads. *Munich Personal RePEc Archive*. https://mpra.ub.uni-muenchen.de/id/eprint/94395.

[Srikanth K. Iyer, 2009] Srikanth K. Iyer, R Emilion, H. F. (2009). Bayesian estimation for markov modulated asset prices.

[Srikanth K. Iyer, 2008] Srikanth K. Iyer, Didier Bernard, R. E. A. T. (2008). A model for dissipation: cascade sde with markov regime-switching and dirichlet prior. ffhal-00286131f. *hal-00286131*.

[Teh, 2010] Teh, Y. W. (2010). Dirichlet process. *University College London*.

[Tierney, 1994] Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*.

[Veraart, 2016] Veraart, A. E. D. (2016). Modelling the impact of wind power production on electricity prices by regime-switching lévy semistationary processes. In *Stochastics of Environmental and Financial Economics*, pages 321–340, Cham. Springer International Publishing.