Available online at http://scik.org
Math. Finance Lett. 2026, 2026:1
https://doi.org/10.28919/mfl/10006
ISSN: 2051-2929

HIDDEN MARKOV MODELLING OF TIME SERIES VARIANCE REGIMES USING LARGE DEVIATION THEORY

TROON JOHN BENEDICT[1,*], RAMKUMARI T. BALAN[2], MUNG'ATU JOSEPH[3], ONYANGO FREDRICK[4]

[1]Department of Economics, Maasai Mara University, Narok, Kenya

[2]Department of Mathematics and Statistics, University of Dodoma, Dodoma, Tanzania

[3]Department of Mathematics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

[4]Department of Mathematics and Actuarial Science, Maseno University, Kisumu, Kenya

Abstract. Heteroskedasticity occurs when the variance of a time series fluctuates. This alters the distribution and modeling of the series. Volatility models such as GARCH and ARCH have demonstrated insufficiency in modeling such series. The shortcomings of the models have led to the development of regime-switching volatility models, which have demonstrated efficacy in modeling such series. Markov switching volatility models exhibit greater efficiency by integrating regime dependencies in the modeling of regime transitions. Nevertheless, the models depend on precise identification and modeling of the fundamental Markov chain that governs the variance regimes. The maximum likelihood estimation (MLE) and expectation maximization (EM) have been used before to model the hidden Markov regimes of such series. This study presents an alternative approach for modeling Hidden Markov Models (HMM) through the Large Deviation Principle (LDP). The approach employed LDP to ascertain variance regimes and MLE to estimate HMM parameters. The Viterbi algorithm decodes the series' variance regimes after estimating the model parameters. We juxtaposed the model results with those of the Expectation Maximization Hidden Markov Model (EM-HMM). The two methods yielded nearly identical sets of variance regimes for heteroscedastic time series, as demonstrated by the simulations and the gold price data. We determined

*Corresponding author

E-mail address: troon@mmarau.ac.ke

Received May 26, 2026

that the model surpassed the EM algorithm in discerning the number of underlying variance regimes. The approach was more effective for modeling the underlying variance regimes when the EM algorithm did not converge. It also proved more effective for modeling the fundamental variance regimes relative to a benchmark variance.

Keywords: heteroscedasticity; variance regime; hidden Markov model; large deviation principle; regime switching; Viterbi algorithm.

2020 AMS Subject Classification: 62M05, 62M10, 60F10.

1. INTRODUCTION

Structural breaks in a time series occur when a policy change or other random factors within the series environment induce alterations in the series' parameters, such as the mean and variance [1].

The breaks adversely impact the modeling of the series by inducing non-stationarity through alterations in its moments. Accurate incorporation of structural breaks is essential for the proper modeling of time series. Moreover, if the forecasting model does not account for the disruptions that arise during the forecasting period, it may undermine the accuracy of the forecast [2].

Variance breaks specifically induce heteroscedasticity in time series data, thereby impacting the modeling of these series. Researchers have developed volatility models, such as ARCH and GARCH, to model heteroscedastic time series [3, 4, 5]. Nevertheless, for series exhibiting variance breaks, these models have proven inadequate because they tend to overestimate or occasionally underestimate the volatility of the series due to enduring effects. This impacts the precision of the model's estimations and predictions [6, 7, 8].

Due to the inadequacy of these volatility models, regime change volatility models, mixture volatility models, and Markov switching volatility models, such as the Markov switching GARCH and ARCH models, were developed. These models have demonstrated superior efficacy in modeling series with variance breaks compared to single regime volatility models over the years [9, 10, 11].

In most time series data exhibiting variance breaks, it is widely recognized that the variance regimes are interdependent rather than independent. Consequently, the Markov switching regimes, as opposed to the independent mixture model regimes, offer the most appropriate method for capturing this dependency. This has resulted in increased popularity of Markov

switching regime models for time series analysis with variance breaks, in contrast to independent mixture models [12, 13, 14].

The accuracy of the models depends on the exact representation of the underlying variance regimes within the series. This entails precise identification of the parameters of the foundational Markov model and the correct categorization of the series observations into the appropriate variance regime [15, 16].

The undelying variance regimes and the series observations constitute a Hidden Markov Model (HMM), wherein the concealed states represent the underlying variance regimes and the emitted/observed states correspond to the series observations. The precision of the Markov switching model depends not only on the accurate representation of the underlying Markov states for variance regimes but also on the thorough modeling of the hidden Markov model [15, 16].

Researchers have devised multiple techniques to model the Hidden Markov Model (HMM), including the Expectation Maximization (EM) method, which optimally aligns the HMM to categorize each sequential observation within the regime of maximum likelihood. The method estimates the model parameters while ensuring the classification of observations into states with the highest probability of occurrence [17, 18]. Another technique employed in HMM modeling is maximum likelihood estimation, which guarantees that the parameter estimates of the HMM are obtained from those with the highest probability of occurrence, thus maximizing the likelihood function [15, 16].

Although the MLE and EM algorithm techniques excel in modeling HMM, their main emphasis is not on capturing the variance regimes of the series, but rather on modeling the regimes that maximize likelihood probability or the probability of an observation belonging to a designated regime. At times, the user's judgment dictates the number of regimes to which the series is assigned, rather than any computational criterion. Consequently, it was essential to devise a technique that concentrated on the HMM of the variance regimes for the time series, while also possessing the ability to objectively ascertain the number of variance regimes for classification of the series. This study presents a novel approach to modeling the variance regimes of time series by initially identifying them through LDP and subsequently employing HMM to model the observed series in conjunction with the identified underlying variance regimes.

1.1. Large Deviation Theory. Large Deviation Theory is a mathematical theory that focuses on the examination and exploration of infrequent occurrences or events that occur at the extreme ends of probability distributions [19]. The theory primarily focuses on the analysis of the probability of the data distribution deviating from the normal distribution and the occurrence of specific rare events. Throughout the years, the theory has found applications in various disciplines such as finance, insurance, queuing concepts, health, mathematical modeling, physics, and engineering, among others [20, 21].

The law of large numbers asserts that as the sample size expands, the sample mean will approach the true population mean. The central limit theorem posits that with a sufficiently large sample size, the distribution of sample means will approximate a normal distribution. The large deviation theory differs marginally from these two theories by concentrating on substantial deviations in the sample mean that are markedly distant from the true population mean. The theory primarily concentrates on controlling the rate at which the probability of observing such deviations escalates or diminishes as the sample size increases. The theory centers on establishing a core principle through partial sums, guaranteeing a swift reduction in the likelihood of diverging from the true mean as the sample size expands. This principle is commonly known as the Large Deviation Principle [22].

1.1.1. *Large Deviation Principle.* Consider a random variable $S_n = X_1 + X_2 + X_3 + ... + X_n$ for $\{X\}_{n\geq 1}$ with a probability density function denoted as $f(S_n)$. The random variable S_n is said to satisfy the large deviation principle if there exists a limit

$$\lim_{n\to\infty} \frac{1}{n} f(S_n = s) = -I(s) \tag{1}$$

where $I(s)$ is the rate function which is non-negative and not everywhere zero.

When the random variable S_n satisfies the large deviation theory, then

$$f(S_n = s) \approx \exp(-nI(s)) \tag{2}$$

This is because

$$f(S_n = s) = \exp(-nI(s) + e(n))$$

where $e(n)$ is the correction term. Hence, when we take the limits as $n \to \infty$

$$\lim_{n\to\infty} -\frac{1}{n} f(S_n = s) = I(s) - \lim_{n\to\infty} \frac{e(n)}{n} = I(s)$$

Hence, Equation 2 demonstrates that the large deviation principle establishes the boundary for preserving the predominant exponential term $f(S_n = s)$ while disregarding any additional sub-exponential terms [23, 22, 24].

1.1.2. *Determination of Rate Function.* [25] [25] states that the primary element of the LDP is the identification of a suitable rate function $I(s)$ that fulfills the LDP. Several methodologies have been devised to ascertain the rate function $I(s)$. This encompasses;

(1) **Direct Method** - It involves the determination of the probability density function for the random variable S_n and then proving that it is in the form of the LDP.

(2) **Indirect Method** - It involves the calculation of another function which is a function of the random variable S_n whose properties can be used in concluding that S_n satisfy the LDP.

One of the techniques in this class that has been developed for deriving the rate function is the **Gartner-Ellis Theorem**. The theorem involves the calculation of the function

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln E(\exp(nkS_n)) \tag{3}$$

which is known as the scaled cumulant generating function. According to the theorem, if Equation 3 is differentiable with respect to k then S_n satisfy the LDP and the rate function is given by

$$I(s) = \sup(ks - \lambda(k)) \tag{4}$$

(3) **Contraction Method** - The method involves relating the random variable S_n to another random variable which is known to satisfy an LDP and then derive from this the LDP for S_n. The principle for doing this is known as **The Contraction Principle**.

1.2. Markov Process. Consider a random variable $\{Z(t), t \geq 0\}$ that varies with the index $t \in T$, where t denotes time. The random variable $Z(t)$ is referred to as a stochastic process in this context [26].

The state space of the random variable $Z(t)$ comprises all potential values that $Z(t)$ may assume, denoted by the set $S = 0, 1, 2, 3, \ldots$. If the transition probability of the random variable $Z(t)$ to state j is exclusively dictated by the current state i and is unaffected by prior states, then $Z(t)$ is characterized by the Markovian property and is consequently classified as a Markov process [27, 28].

The one-step transition probability, represented as p_{ij}, indicates the likelihood of moving from state i to state j in a Markov process $Z(t)$. This indicates the likelihood of moving from state i to state j in one step. The process $Z(t)$ generally possesses a one-step transition probability matrix P that consolidates the transition probabilities to all states. All elements in matrix P are generally non-negative and possess a maximum value of 1. The aggregate of the values in each row of the matrix generally totals 1 [29].

Markov processes have recently gained prominence across diverse application domains, especially in forecasting and predictive modeling. These processes are especially advantageous in scenarios where forthcoming observations are significantly affected by the present condition, rather than by occurrences from the remote past. The method has been employed in diverse applications including time series forecasting, market trend analysis, digital imaging, speech signal processing, and queueing processes, among others. The process has enabled the classification of real-world events into interconnected states, allowing for the assessment of the probability of transitioning between these states. As a result, predicting and forecasting has become more straightforward [30].

1.3. Hidden Markov Model. Examine a stochastic process represented by $X(t)$. A process $X(t)$ is designated as a Hidden Markov Model (HMM) if its generating distribution is contingent upon the states of an unobserved Markov process. The concealed process that produces $X(t)$ is defined by a Markov process with latent states that conform to the Markovian property. The process $X(t)$ transitions between hidden states, governed by a transition matrix, akin to a conventional Markov chain. The observed values of the process $X(t)$ are termed the emission states and are linked to the latent states of the process via an emission matrix. The concealed states of the process $X(t)$ demonstrate correlation with one another [31].

Acoording to [15] [15], a hidden Markov process is governed by three key distributions.

(1) **Initial Distribution** - This illustrates the allocation of the initial concealed states of the process. The distribution is sometimes presumed to be stationary or non-stationary. Generally, the concealed states are discrete, thus the initial distribution is frequently depicted by probability mass functions such as Bernoulli, Multinomial, Binomial, and Poisson.

(2) **Transition Distribution** - This distribution signifies the transition among various concealed states of the process. When the hidden states are discrete, the transition distribution constitutes a probability mass function.

(3) **Emission Distribution** - This illustrates the distribution of the observed states of the process, considering the hidden states. The distribution offers the conditional probability of observing a specific value of the process, contingent upon its presence in a designated hidden state. The distribution depends on the type of observed variable. If the observed variable is continuous, its distribution is characterized by a probability density function. Nonetheless, if the variable is discrete or categorical, its distribution is depicted by a probability mass function.

2. METHODOLOGY

2.1. Large Deviation Principle. Consider $X(t)$ as a time series variable that follows a normal distribution with a mean of μ_t and a variance of σ_t^2. Let us examine a subtime interval $\delta t \in \mathbb{N}^+$ and $1 < \delta t < t$ where t is greater than 2. The subsection variance $V(\delta t)$ in the sub-time interval δt is defined as

$$V(\delta t) = \frac{\sum (X_{\delta t} - \bar{X_{\delta t}})^2}{\delta t - 1} \tag{5}$$

where $\bar{X_{\delta t}}$ is the average value of the series at a specific sub time interval δt.

Consider the subsections standardized variance $v_{\delta t}$ which compares the subsection variance $V(\delta t)$ and the series variance σ_t^2 defined as

$$v_{\delta t} = \frac{(\delta t - 1)V(\delta t)}{\sigma_t^2}$$

We aim to estimate the probability $Pr(v_{\delta t} > a)$, where a is a fixed threshold for the standardised variance $v_{\delta t}$. The probability can be evaluated via the LDP function in the following manner

$$Pr(v_{\delta t} > a) \approx \exp\left(-\delta t(I(a))\right)$$

where $I(a)$ represents the rate function. The rate function $I(a)$ was obtained through the use of the Gartner Ellis theorem.

Using the approximated probability, a significance level α the study used indicator variable $I_b(\delta t)$ to identify sections of variance variations.

$$I_b(t) = \begin{cases} 0 & Pr(v_{\delta t} > a) \geq \alpha \\ 1 & Pr(v_{\delta t} > a) < \alpha \end{cases}$$

The value of $I_b(\delta t) = 1$, when the series $X(t)$ undergoes a variance change in subsection δt and $I_b(\delta t) = 0$, provided the series does not undergo a change variance of magnitude a in the sub-time interval δt.

2.2. Hidden Markov Model. After identifying the m variance phases of the series resulting from the variance alteration, the Hidden Markov Model (HMM) is utilized to model the time series $X(t)$.

Examine a time series model $X(t)$ comprising m phases. Let $S = \{1,2,3,\ldots,m\}$ represent the set of all potential phases of the series. The transition of the series across various states influences the value of the series $X(t)$ at time t. Consequently, the Markov Model is employed to represent the variance phases of the series as hidden or unobserved states, whereas the actual values of the series $X(t)$ are modeled as the emission states of the Hidden Markov Model (HMM).

2.2.1. *Initial Distribution.* It is assumed that the time series $X(t)$ comprises m phases, regarded as the hidden states of the Hidden Markov Model (HMM), and these phases are presumed to be discrete due to the anticipated countable number of interruptions in the series $X(t)$. The initial distribution $\pi_i(1)$ defines the probability distribution of each variance phase of the series $X(t)$ at time $t = 1$.

2.2.2. *Transition Distribution.* Let $Z(t)$ represent a random variable indicating the phase of the series $X(t)$ at time t, where $Z(t)$ assumes values from the set $\{1,2,...,m\}$. The distribution of $Z(t)|Z(t-1)$ is presumed to follow a multinomial distribution with a transition probability consistent with the Markov property such that

$$Pr(Z(t)|Z(1),Z(2),Z(3),...,Z(t-1)) = Pr(Z(t)|Z(t-1)). \tag{6}$$

The assumption of a multinomial distribution is predicated on the existence of m distinct and non-overlapping phases in the series, which are considered countable.

The transition probabilities among phases in the series are consolidated into a matrix P, known as the transition probability matrix. The elements p_{ij} denote the probability of transitioning from phase i to phase j in a single step at any specified time t. Each row of matrix P sums to 1.

2.2.3. *Emission Distribution.* The output values of the Hidden Markov Model (HMM) correspond to the values of the time series $X(t)$ at a specific time t, which is contingent upon the variance regimes at that time $Z(t)$. The probability distribution of the random variable $X(t)|Z(t)$ is presumed to be approximately normal, characterized by a mean μ_i and a variance σ_i^2.

3. MATHEMATICAL MODELLING

3.1. Large Deviation Principle For Subsection Variance Break. Examine the time series $X(t)$ for $t > 0$. The series is divided into equal segments of size δt, where $\delta t \geq 10$, and the total number of segments is no more than 30. If the series $X(t)$ comprises T time points, then $\frac{T}{\delta t} \leq 30$. The subsection size δt is established at a minimum of 10, as a sample size below this threshold renders the variance less dependable due to its heightened sensitivity to outliers. Small data sets are characterized by a less symmetrical distribution, resulting in skewness that impacts variance estimation [32]. The limit of 30 subsections is established to avert an excessive quantity, as a sample size exceeding 30 is deemed substantial. Each subsection δt is represented by its local variance, denoted as $V(\delta t)$. Our objective is to compare the variance of each subsection with that of the overall series.

Denote the variance of the complete series as σ_t^2 at time $t > 0$. In a homoscedastic series, the variance of each subsection $V(\delta t)$ is expected to be uniform with the variance of the entire series. If heteroscedasticity is present, the variance of each subsection will differ from the

overall series variance. Define $v_{\delta t} = \frac{(\delta t-1)V(\delta t)}{\sigma_t^2}$ as the ratio of the subsection variance to the total variance of the series.

Let a denote a positive threshold for the variance ratio to indicate a significant disparity between the subsection variance and the overall series variance. Our objective is to obtain an accurate estimate of $Pr(v_{\delta t} \geq a)$ to determine whether the variance of the subsection significantly differs from the variance of the series. The LDP rate function for estimating $Pr(v_{\delta t} \geq a)$ was established by [33] [33] to be expressed by Equation 7;

$$I(a) = \frac{a}{2\delta t} - \frac{1}{2} + \frac{1}{2}\ln(\frac{\delta t}{a}) \tag{7}$$

The full derivation of this rate function can be obtained in [33] and also shown in appendix A. The rate function is also established to be able to estimate $Pr(a < v_{\delta t})$ for the threshold value $a < v_{\delta t}$, where $a \geq 0$ [33].

3.2. Variance Classification Rule. Consider a series $X(t)$, a significance level $\alpha = 0.001$, a rate function $I(a)$ defined in Equation 7, a local subsection size δt, a local series variance $V(\delta t)$, and a standardized local series variance $v_{\delta t}$. If the LDP probability, as determined by the rate function $I(a)$, is lower than the threshold $\alpha = 0.001$, and the local standardized variance $v_{\delta t}$ is less than the subsection size δt, then the series $X(t)$ is deemed to significantly decrease variance from moderate or normal levels to a substantially lower level. On the other hand, if the LDP probability is lower than $\alpha = 0.001$ and the standardized local variance $v_{\delta t}$ is greater than the subsection size δt, the series $X(t)$ is considered to exhibit a significant increase in variance. This can be represented using the indicator function $I(V(\delta t))$ as follows [34]:

$$I(V(\delta t)) = \begin{cases} -1 & \text{if} \quad Pr(v_{\delta t} \geq a) < \alpha \quad \text{and} \quad v_{\delta t} < \delta t \\ 0 & \text{if} \quad Pr(v_{\delta t} \geq a) \geq \alpha \\ 1 & \text{if} \quad Pr(v_{\delta} \geq a) < \alpha \quad \text{and} \quad v_{\delta t} > \delta t \end{cases} \tag{8}$$

The value $I(V(\delta t)) = -1$ indicates a considerably low variance, $I(V(\delta t)) = 0$ indicates normal or average variance, and $I(V(\delta t)) = 1$ indicates a very high variance.

3.3. Definition of the Hidden Markov Model.

3.3.1. *Hidden Layer.* Let $X(t)$ represent a multiphased time series characterized by variance phases as illustrated by the LDP defined in the previous section. Let the stochastic process $Z(t)$ represent the underlying variance regime of the series $X(t)$ at time $t > 0$. $Z(t)$ can take on the values $1,...,m$, with each value representing the corresponding variance regime of the series $X(t)$ at time t. Assuming the variance regimes serve as the underlying hidden states of the series $X(t)$, we posit that the process $Z(t)$ adheres to the Markov property such that

$$Pr(Z(t)|Z(1),Z(2),Z(3),...,Z(t-1)) = Pr(Z(t)|Z(t-1)) \tag{9}$$

3.3.2. *Transition, Initial and Stationary Distribution of $Z(t)$.* The aggregated one-step chance of transition of the process $Z(t)$ between the m states is given by the transition probability matrix P where

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix}$$

and $p_{ij} \geq 0$ denotes the probability of the transition from state $i \in m$ to $j \in m$ in a single step. The process $Z(t)$ in this case is considered to be a Markov process with transition probability matrix P.

At time $t = 1$, the process $Z(t)$ is assumed to be multinomial distributed with probability mass function

$$f(Z_i) = \left\{ \begin{array}{ll} \frac{t!}{Z_1!...Z_m!}\pi_1^{Z_1}...\pi_m^{Z_m} & i = 1,...,m \\ 0 & \text{otherwise} \end{array} \right\}$$

where Z_i is the number of times the series is in state i and π_i is the probability that the series is in state i at time $t = 1$. Assuming that the process $Z(t)$ homogenous, the probability of the series transitioning to a given states at time $t > 0$ is given by the t step transition matrix P^t. Hence, the distribution of the series at time $t > 0$ will be given by

$$\pi(t) = \pi(1)P^{t-1} \tag{10}$$

where $\pi(t)$ is $1 \times m$ vector of probabilities, showing the distribution of the process $Z(t)$ at time $t > 0$. $\pi(1)$ is $1 \times m$ vector of probabilities, showing the initial distribution of the process $Z(t)$ at time $t = 1$ and P^{t-1} is the $t-1$ steps transition probability matrix for the process $Z(t)$.

At some point $t^* > 0$, the process will be assumed to attain a stationary distribution π such that

$$\pi P^{t-1} = \pi \tag{11}$$

At this point (t^*) the process $Z(t)$ is considered to be a stationary process.

Let P be the one step transition probability matrix for the m state hidden layer $Z(t)$ of the hidden markov model $X(t)$, where each entry p_{ij} represents the probability of moving from state i to state j in a single step

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}.$$

The entries satisfy $0 \leq p_{ij} \leq 1$ and $\sum_{j=1}^{n} p_{ij} = 1$ for each i.

Let $\pi = (\pi_1, \ldots, \pi_m)$ be the stationary distribution vector, where π_i represents the long-term probability of being in state i. The vector π is required to be such that $\pi P = \pi$ and $\sum_{i=1}^{m} \pi_i = 1$. In order to determine the values of π that satisfy these properties, the following computation was done on the transition matrix P to determine the required values of π.

The first property of π requires that $\pi P = \pi$. This can be expanded as

$$\begin{aligned} \pi_1 &= \pi_1 p_{11} + \pi_2 p_{21} + \cdots + \pi_m p_{m1} \\ \pi_2 &= \pi_1 p_{12} + \pi_2 p_{22} + \cdots + \pi_m p_{m2} \\ &\vdots \\ \pi_m &= \pi_1 p_{1m} + \pi_2 p_{2m} + \cdots + \pi_m p_{mm}. \end{aligned}$$

This can be written in matrix form,as

$$\pi(P - I) = 0, \tag{12}$$

where I is an $m \times m$ identity matrix. The equation 12 represents a homogeneous system of linear equations

$$(p_{11}-1)\pi_1 + p_{21}\pi_2 + \cdots + p_{m1}\pi_m = 0$$

$$p_{12}\pi_1 + (p_{22}-1)\pi_2 + \cdots + p_{m2}\pi_m = 0$$

$$\vdots$$

$$p_{1m}\pi_1 + p_{2m}\pi_2 + \cdots + (p_{mm}-1)\pi_m = 0.$$

To obtain a unique solution for these systems of linear equations we add the normalization condition based on the second property of π such that

$$\sum_{i=1}^{m} \pi_i = 1.$$

This gives us an augmented system

$$\begin{pmatrix} & P-I & \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Solving this system of linear equations using row reduction will yield the stationary distribution vector $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$.

3.3.3. *Emission Layer.* The process $X(t)$ is considered as a hidden markov process with hidden states $Z(t)$ at time $t > 0$. It follows that the distribution of $X(t)$ at time $t > 0$ depends on the state $Z(t)$ at time $t > 0$ such that

$$Pr(X(t)|X(1),...,X(t-1),Z(1),...,Z(t)) = Pr(X(t)|Z(t)) \tag{13}$$

Given that the series $X(t)$ is a continuous series. In each of the states $i = 1,...,m$, the series $X(t)$ is assumed to be normally distributed with mean μ_i and variance σ_i^2. Therefore, if $f_i(x)$ denotes the distribution of the series in state i, then

$$f_i(x) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(x-\mu_i)^2\right) & -\infty \leq x \leq \infty \\ 0 & \text{otherwise} \end{array} \right\} \tag{14}$$

and

$$Pr(X(t) = x|Z(t) = i) = f_i(x) \tag{15}$$

where $Pr(X(t) = x|Z(t) = i)$ is the probability of observing a value x for the series $X(t)$ at time $t > 0$ given that the series $X(t)$ is in state i.

3.4. Distribution of Resulting Hidden Markov Model ($X(t)$). Let $\pi_i(t) = Pr(Z(t) = i)$, we wish to determine a formulation for $f(X(t) = x)$ which is the distribution $X(t)$ at time $t > 0$. By definition,

$$\begin{aligned} f(X(t) = x) &= \sum_{i=1}^{m} Pr(X(t) = x|X(1), ..., X(t-1), Z(1), ..., Z(t) = i) \\ &= \sum_{i=1}^{m} Pr(Z(t) = i) Pr(X(t) = x|Z(t) = i) \\ &= \sum_{i=1}^{m} \pi_i(t) f_i(x) \end{aligned}$$

This can be written in matrix form as

$$f(X(t) = x) = \pi(t) F(x) 1^{'} \tag{16}$$

where

- $\pi(t) = (\pi_1(t), ..., \pi_m(t))$ is a $1 \times m$ vector of $Pr(Z(t) = i)$ for $i = 1, ..., m$
- $F(x) = \begin{bmatrix} f_1(x) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & f_m(x) \end{bmatrix}$ is an $m \times m$ diagonal matrix with the main diagonal representing the probability density function $f_i(x)$ for $i = 1, ..., m$
- 1 is a $1 \times m$ vector of $1s$.

Given that P is a one step transition matrix for the random variable $Z(t)$ as earlier defined and $\pi(1)$ is the distribution vector for $Pr(Z(1) = i)$. Then, we can write the distribution function for $X(t) = x$ as

$$f(X(t) = x) = \pi(1) P^{t-1} F(x) 1^{'} \tag{17}$$

Now, if we consider the situation in which $Z(t)$ is stationary with stationary distribution vector π. Then, in this case

$$f(X(t) = x) = \pi F(x) 1^{'} \tag{18}$$

due to the fact that $\pi P^{t-1} = \pi$.

3.5. Parameter Estimation. Let $x(1),x(2),x(3),\ldots,x(t)$ represent the observations of the time series $X(t)$ from time $T = 1$ to time $T = t$. The joint distribution for this sequence of observations is represented as follows:

$$f(x(1),x(2),x(3),...,x(t)) = \pi(1)F(x(1))PF(x(2))PF(x(3))...PF(x(t))1' \tag{19}$$

The joint distribution in this instance is synonymous with the likelihood function for the random variables, which is also expressed as

$$L(x(1),x(2),x(3),...,x(t)) = \pi(1)F(x(1))PF(x(2))PF(x(3))...PF(x(t))1'$$

This indicates that, to estimate the parameters via the maximum likelihood estimation technique, it is necessary to identify the parameter estimates that will maximize the likelihood function. The LDP technique facilitated the decoding of the hidden state for the series $X(t)$, which are influenced by the underlying variance regime. Consequently, this indicates that the observations of the series within each regime are predetermined. This streamlines the MLE estimation by removing the necessity to maximize the MLE function in the presence of missing observations (underlying Markov observations), thereby obviating the need for the EM algorithm which involves estimating the model parameters while determining the possible underlying states that will optimize the likelihood function. The optimization of the current likelihood function can be achieved by maximizing its individual components: initial distribution parameters, transition matrix parameters, and emission distribution parameters. Consequently, the maximum likelihood estimation (MLE) for the model parameters entailed the MLE estimation for the parameters of the hidden state. Then, using the hidden state parameters, determine the emission distribution parameters that will optimize the log likelihood function.

3.5.1. *Initial Distribution.* The multiphased series $X(t)$ is represented as a hidden Markov model with latent states $Z(t)$, which correspond to the variance phases generated by the LDP described in chapter 4. The LDP technique permitted the categorization of all observations of $X(t)$ into variance phases. This signifies that the LDP enabled the determination of all the $Z(t)$ values. This signifies that each $X(t)$ can be categorized as a $Z(t)$ at time t.

Consider the phases of the series $X(t)$ at time $t = 1$, assuming that the distribution of the m variance phases at this time follows a multinomial distribution with parameters π_i. The observations

$Z_1, Z_2, \ldots, Z_m$ denote the frequency of occurrences in each of the m variance phases, with the probability of occurrence in the i^{th} phase represented by π_i, where $\pi_i \geq 0$ and $\sum_{i=1}^{m} \pi_i = 1$. The joint probability mass function for $Z_1, Z_2, \ldots, Z_m$ is defined by the multinomial distribution

$$P(Z_1 = z_1, Z_2 = z_2, \ldots, Z_m = z_m) = \frac{t!}{z_1! z_2! \cdots z_m!} \pi_1^{z_1} \pi_2^{z_2} \cdots \pi_m^{z_m}$$

where $z_1 + z_2 + \cdots + z_m = t$ and t is the total number of observation of the series $X(t)$.

At time $t = 1$, the series $X(t)$ is observed to be in state $Z(1) = i$, indicating that the state of the series evolves with time. However, subsequently transition to alternative states $Z(t)$ at periods $t > 1$. At time $t = 1$, we are unaware of any alternative variance regime or state into which the series may transition. Consequently, based on the knowledge known at time $t = 1$, the series will undoubtedly be in state $Z(1) = i$. Consequently, the initial distribution is configured so that the state where $X(1)$ starts at time $t = 1$ ($Z(1) = i$) has a probability of 1, while all other states possess a probability of 0. This represents the estimate of the initial distribution $\hat{\pi}(1)$.

3.5.2. *Transition Probability Matrix.* Let p_{ij} represent the transition probability from variance phase i to phase j in a single time step, and let f_{ij} express the frequency of transitions from state i to state j in a single step.

Considering that the sum of each row of P equals 1. The matrix has $m(m-1)$ free parameters, as $p_{ii} = 1 - \sum_{k \neq i} p_{ik}$. We aim to estimate each of the $m(m-1)$ parameters utilising Maximum Likelihood Estimation (MLE). The joint distribution of the latent states $Z(1), Z(2), ..., Z(t)$ is expressed as

$$f(Z(1), Z(2), ..., Z(t)) = \Pi_{i=1}^{m} \Pi_{j=1}^{m} p_{ij}^{f_{ij}}$$

This shows that the likelihood function is

$$L(Z(t), p_{ij}) = \Pi_{i=1}^{m} \Pi_{j=1}^{m} p_{ij}^{f_{ij}}$$

Therefore, the log-likelihood function is

$$\ln L(Z(t), p_{ij}) = \sum_{i=1}^{m} (\sum_{j=1}^{m} f_{ij} \ln p_{ij})$$

$$= \sum_{i=1}^{m} l_i$$

We can maximize $\ln L(Z(t), p_{ij})$ by maximizing each of the l_i. But we know that

$$l_i = f_{ii} \ln (1 - \sum_{j \neq i} p_{ij}) + f_{ij} \ln (p_{ij})$$

Therefore, if we differentiate l_i with respect to p_{ij} and equate to 0

$$\begin{aligned} \frac{\partial l_i}{\partial p_{ij}} &= \frac{-f_{ii}}{1 - \sum_{j \neq i} p_{ij}} + \frac{f_{ij}}{p_{ij}} \\ &= \frac{-f_{ii}}{p_{ii}} + \frac{f_{ij}}{p_{ij}} = 0 \\ \frac{f_{ii}}{p_{ii}} &= \frac{f_{ij}}{p_{ij}} \\ f_{ii} p_{ij} &= p_{ii} f_{ij} \end{aligned}$$

summing both sides of the resulting equation with respect to j

$$\begin{aligned} \sum_{j=1}^{m} f_{ii} p_{ij} &= \sum_{j=1}^{m} p_{ii} f_{ij} \\ f_{ii} \sum_{j=1}^{m} p_{ij} &= p_{ii} \sum_{j=1}^{m} f_{ij} \\ f_{ii} &= p_{ii} \sum_{j=1}^{m} f_{ij} \end{aligned}$$

This is due to the fact that $\sum_{j=1}^{m} p_{ij} = 1$. This shows that the MLE estimator for the main diagonal value of the transition matrix p_{ii} will be given by

$$\hat{p}_{ii} = \frac{f_{ii}}{\sum_{j=1}^{m} f_{ij}} \tag{20}$$

and for the off diagonal values p_{ij}

$$\hat{p}_{ij} = \frac{f_{ij} p_{ii}}{f_{ii}} \tag{21}$$

but $\frac{p_{ii}}{f_{ii}} = \frac{1}{\sum_{j=1}^{m} f_{ij}}$. Therefore,

$$\hat{p}_{ij} = \frac{f_{ij}}{\sum_{j=1}^{m} f_{ij}} \tag{22}$$

In this case, the MLE estimator for the transition matrix P is

$$\hat{P} = \begin{bmatrix} \hat{p}_{11} & \cdots & \hat{p}_{1m} \\ \vdots & \ddots & \vdots \\ \hat{p}_{m1} & \cdots & \hat{p}_{mm} \end{bmatrix} \tag{23}$$

3.5.3. *Emission Distribution.* In each state $i = 1, ..., m$, the emission distribution $F(x_i)$ is presumed to follow a normal distribution characterized by mean μ_i and variance σ_i^2. The probability density function for the random variable x_i in each state i is expressed as follows

$$f(x_i(t) = x_k) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(x_k - \mu_i)^2\right) & -\infty \leq x_k \leq \infty \\ 0 & \text{otherwise} \end{array} \right\}$$

Our objective is to ascertain the parameter estimates for the probability density function for each state that will optimise the likelihood function, conditioned upon the estimations of the hidden state distribution ($\hat{\pi}(1)$ and $\hat{P}$).

Let $X_1, X_2, \ldots, X_{n_i}$ be the observations of the series $X(t)$ which are classisfied by the LDP technique to be in the variance regime i. For the variance regime i, the observations $X_1, X_2, \ldots, X_{n_i}$ are assumed to be normally distributed with mean μ_i and variance σ_i^2.

The likelihood function $L(\mu_i, \sigma_i^2 \mid X_1, X_2, \ldots, X_{n_i})$ for the variables is therefore given by:

$$L(\mu_i, \sigma_i^2 \mid X_1, X_2, \ldots, X_{n_i}) = \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(X_j - \mu_i)^2}{2\sigma_i^2}\right)$$

As a result, the log-likelihood function $\ell(\mu_i, \sigma_i^2)$ is given by:

$$\ell(\mu_i, \sigma_i^2) = \sum_{j=1}^{n_i} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) - \frac{(X_j - \mu_i)^2}{2\sigma_i^2}$$

$$\ell(\mu_i, \sigma_i^2) = -\frac{n_i}{2}\log(2\pi\sigma_i^2) - \frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}(X_j - \mu_i)^2$$

To find the MLE for μ_i, we take the derivative of the log-likelihood function with respect to μ_i and set it to zero:

$$\frac{\partial \ell}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}\left(-\frac{n_i}{2}\log(2\pi\sigma_i^2) - \frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}(X_j - \mu_i)^2\right)$$

$$\frac{\partial \ell}{\partial \mu_i} = -\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (X_j - \mu_i) = 0$$

$$\sum_{j=1}^{n_i} X_j - n_i \mu_i = 0$$

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j$$

Thus, the MLE for μ_i is:

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j = \bar{X}_i$$

Next, we find the MLE for σ_i^2 by taking the derivative of the log-likelihood function with respect to σ_i^2 and setting it to zero:

$$\frac{\partial \ell}{\partial \sigma_i^2} = \frac{\partial}{\partial \sigma_i^2} \left(-\frac{n_i}{2} \log(2\pi\sigma_i^2) - \frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} (X_j - \mu_i)^2 \right)$$

$$\frac{\partial \ell}{\partial \sigma_i^2} = -\frac{n_i}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} \sum_{j=1}^{n_i} (X_j - \mu_i)^2 = 0$$

$$-\frac{n_i \sigma_i^2}{2\sigma_i^4} + \frac{1}{2\sigma_i^4} \sum_{j=1}^{n_i} (X_j - \mu_i)^2 = 0$$

$$-\frac{n_i}{2\sigma_i^2} + \frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} (X_j - \mu_i)^2 = 0$$

$$n_i \sigma_i^2 = \sum_{j=1}^{n_i} (X_j - \mu_i)^2$$

$$\sigma_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_j - \mu_i)^2$$

Now we apply the Bessel correction, by adjusting the denominator from n_i to $n_i - 1$:

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_j - \hat{\mu}_i)^2$$

The standard errors of the estimators can be found using the Fisher Information Matrix. For μ_i, the standard error is:

$$\mathrm{SE}(\hat{\mu}_i) = \frac{\sigma_i}{\sqrt{n_i}}$$

For σ_i^2 with Bessel's correction, the standard error is:

$$\mathrm{SE}(\hat{\sigma}_i^2) = \frac{\hat{\sigma}_i^2 \sqrt{2}}{(n_i - 1)}$$

3.6. Length of Stay and Transition.

3.6.1. *Length of Stay.* Let $Z(t) = k$ be the variance phase of the series $X(t)$ at time $t > 0$. We are interested in the period of time that the series will remain in the variance phase k before moving to another variance phase. The length of time that the series $X(t)$ stays in the variance phase k is equivalent to the time until the series switches to any of the other variance phases $m-1$.

Let $\mathscr{A}$ be the set of all possible variance phases that the series $X(t)$ can switch to other than k. The expected time until the series switch to variance phase set $\mathscr{A}$ from variance phase k is given by

$$S_{k\mathscr{A}} = \left\{ \begin{array}{ll} 0 & k \in \mathscr{A} \\ 1 + \sum_{i \notin \mathscr{A}} p_{ki} S_{iA} & k \notin \mathscr{A} \end{array} \right\}$$

where $S_{k\mathscr{A}}$ is the expected time until the series switch from variance phase k to $\mathscr{A}$.

Given that the series can assume m variance phases, the length of stay in variance phase k for the series $X(t)$ is in this case given by

$$\begin{aligned} S_k &= S_{k\mathscr{A}} \\ &= 1 + p_{kk} S_{k\mathscr{A}} + \sum_{i \in \mathscr{A}} p_{ki} S_{iA} \\ &= 1 + p_{kk} S_{k\mathscr{A}} \\ &= 1 + p_{kk} S_k \\ S_k - p_{kk} S_k &= 1 \\ S_k (1 - p_{kk}) &= 1 \\ S_k &= \frac{1}{1 - p_{kk}} \end{aligned}$$

This shows that the expected time for series $X(t)$ to stay in variance phase k given that it enters the variance phase k at time t is given by

$$S_k = \frac{1}{1 - p_{kk}} \tag{24}$$

3.6.2. *Time to switch to specific state.* Consider the case in which the series $X(t)$ enters the variance phase k at time t. Suppose that we are interested in the time it will take for the series to hit the variance phase $i \neq k$. The expected time required for the series to hit state i from k which is denoted by S_{ki}, can be derived for the case in which $m = 2$ or $m = 3$ as follows.

3.6.3. *When $m = 2$.* By definition, the expected time for the series to switch to variance phase i from k is given as

$$S_{ki} = \left\{ \begin{array}{ll} 0 & k = j \\ 1 + \sum_{j \neq i} p_{kj} S_{ji} & k \neq j \end{array} \right\}$$

This shows that, the expected time for the series to switch from variance phase k to i for a two phase series is given by

$$S_{ki} = 1 + p_{kk} S_{ki} + p_{ki} S_{ii}$$

but $S_{ii} = 0$. Therefore

$$S_{ki} = 1 + p_{kk} S_{ki}$$

$$S_{ki} - p_{kk} S_{ki} = 1$$

$$S_{ki} = \frac{1}{1 - p_{kk}}$$

This shows that, for a series with two variance phases k and i, the expected time taken for the series to switch from variance phase k to i is given by

$$S_{ki} = \frac{1}{1 - p_{kk}} \tag{25}$$

which is equivalent to the expected time of staying in variance phase k. Therefore, $S_k = S_{ki}$ for a two phased variance regime series $X(t)$.

3.6.4. *When $m = 3$.* Consider the case in which the series $X(t)$ has three variance phases (k, i, j). The expected type for the series to hit phase i from phase k is given as

$$S_{ki} = 1 + p_{kk} S_{ki} + p_{ki} S_{ii} + p_{kj} S_{ji}$$

$$= 1 + p_{kk} S_{ki} + p_{kj} S_{ji}$$

but

$$\begin{aligned} S_{ji} &= 1 + p_{jj}S_{ji} + p_{jk}S_{ki} + p_{ji}S_{ii} \\ &= 1 + p_{jj}S_{ji} + p_{jk}S_{ki} \\ S_{ji} - p_{jj}S_{ji} &= 1 + p_{jk}S_{ki} \\ S_{ji} &= \frac{1 + p_{jk}S_{ki}}{1 - p_{jj}} \\ &= \frac{1}{1 - p_{jj}} + \frac{p_{jk}}{1 - p_{jj}}S_{ki} \end{aligned}$$

Therefore,

$$\begin{aligned} S_{ki} &= 1 + p_{kk}S_{ki} + p_{kj}S_{ji} \\ &= 1 + p_{kk}S_{ki} + p_{kj}\left(\frac{1}{1 - p_{jj}} + \frac{p_{jk}}{1 - p_{jj}}S_{ki}\right) \\ &= 1 + p_{kk}S_{ki} + \frac{p_{kj}}{1 - p_{jj}} + \frac{p_{kj}p_{jk}}{1 - p_{jj}}S_{ki} \end{aligned}$$

Now, solving for S_{ki}

$$\begin{aligned} S_{ki} &= 1 + p_{kk}S_{ki} + \frac{p_{kj}}{1 - p_{jj}} + \frac{p_{kj}p_{jk}}{1 - p_{jj}}S_{ki} \\ S_{ki} - p_{kk}S_{ki} - \frac{p_{kj}p_{jk}}{1 - p_{jj}}S_{ki} &= 1 + \frac{p_{kj}}{1 - p_{jj}} \\ S_{ki}\left(1 - p_{kk} - \frac{p_{kj}p_{jk}}{1 - p_{jj}}\right) &= 1 + \frac{p_{kj}}{1 - p_{jj}} \\ S_{ki}\left(\frac{1 - p_{jj} - p_{kk} + p_{jj}p_{kk} - p_{kj}p_{jk}}{1 - p_{jj}}\right) &= \frac{1 - p_{jj} + p_{kj}}{1 - p_{jj}} \\ S_{ki} &= \frac{1 - p_{jj} + p_{kj}}{1 - p_{jj} - p_{kk} + p_{jj}p_{kk} - p_{kj}p_{jk}} \end{aligned}$$

This shows that for a series with three variance phases (k, i, j), the expected time until the series hit state i given that it arrives in state k at time t is given by

$$S_{ki} = \frac{1 - p_{jj} + p_{kj}}{1 - p_{jj} - p_{kk} + p_{jj}p_{kk} - p_{kj}p_{jk}} \tag{26}$$

4. Application to Simulated Data

4.1. Discrete Series with Mean and Variance Break.

4.1.1. *Simulation Process.* We created a sequence of 180 observations from a Poisson distribution with varying λ values over time. For the initial 60 observations, $\lambda = 75$; for the subsequent 60 observations, $\lambda = 0.6$; and for the final 60 observations, $\lambda = 7$. Each subsection's simulated value was augmented or randomized by incorporating normal random noise from a normal distribution with a mean of 1 and a standard deviation of 0.3, utilizing the *rnorm* function in R. The fluctuations in λ resulted in alterations to the mean and variance of the series over time, whereas the random error introduced noise into the series. The R code for the simulation is presented in Appendix B.

4.1.2. *Resulting Series.* The resulting series $X(t)$ from the simulation was as illustrated in Figure 1

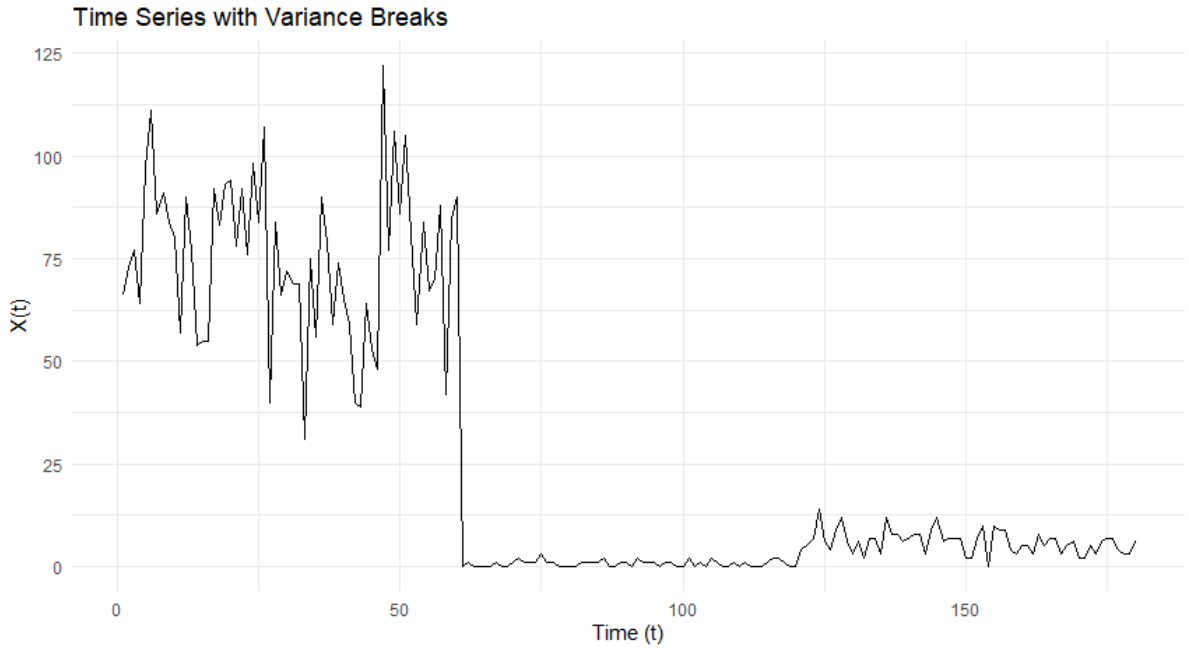


FIGURE 1. Stationary Heteroscedastic Series

The observations in Figure 1 shows a significant fluctuation in the variance and mean of the series over time.

4.1.3. *Hidden Markov Modelling.* We analyzed the variance regimes of the series utilizing the LDP methodology and the EM algorithm. The LDP technique employed a subsection size of

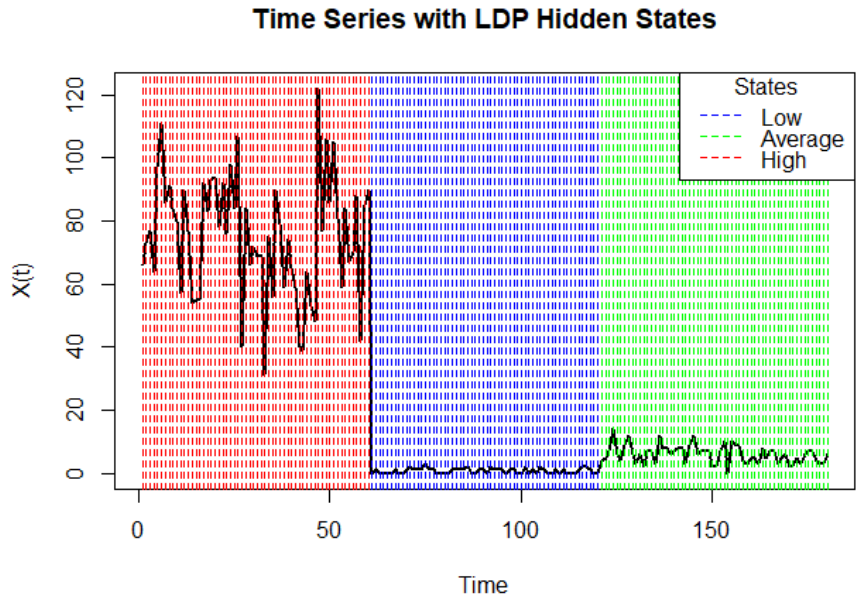
$\delta t = 15$ and a contrasting variance of $\sigma^2 = 10$. The resultant HMMs with parameter estimates are presented in Table 1.

TABLE 1. Parameter Estimates for Hidden Markov Models

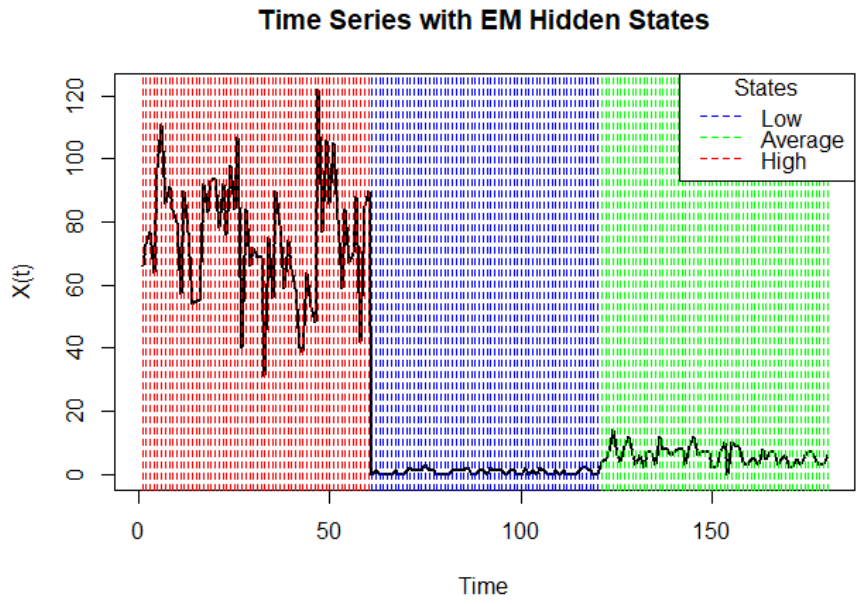
Parameter	LDP Estimate	EM Estimate
p_{11}	0.983	0.983
p_{12}	0.017	0.017
p_{22}	1.000	1.000
p_{23}	0.000	0.000
p_{31}	0.017	0.017
p_{33}	0.983	0.983
$\pi_1(1)$	0	0
$\pi_2(1)$	0	0
$\pi_3(1)$	1	1
μ_1	0.700	0.701
σ_1	0.743	0.737
μ_2	6.017	6.012
σ_2	2.885	2.865
μ_3	75.133	75.133
σ_3	19.289	19.127
π_1	0	0
π_2	1	1
π_3	0	0
Log-link	-679.1479	-487.3618
AIC	1386.296	1002.724
BIC	1430.997	1047.425

The results in table 1 showed that the EM algorithm was a better fit to the data compared to the LDP technique because it had the smallest AIC and BIC. However, the model parameter estimates were remarkably almost similar. This demonstrates that the effectiveness of the LDP

technique was almost equivalent to that of the EM algorithm. Figures 2a and 2b illustrate the graphical representation of the decoded variance regimes for both Hidden Markov Models utilizing the Viterbi Algorithm.



(A) Series with LDP Technique States



(B) Series with EM Algorithm States

The results in Figures 2a and 2b illustrate the similarity in the concealed decoded states for the series $X(t)$ generated by the two algorithms. This indicates that both methods yield identical outcomes.

4.2. Simulated Series from Skewed Distribution.

4.2.1. *Simulation Process.* A chi-square distribution with varying degrees of freedom was employed to simulate 150 observations over time. The degrees of freedom fluctuated from 1.2 for the initial 25 observations to 100 for the subsequent 25, then from 8 for the next 25 to 0.9, followed by 120 for the next 25 to 12, and so forth. The *rnorm* function in R was utilized to introduce normal random noise from a normal distribution with a mean of 1 and a standard deviation of 0.3 to each subsection's simulated value, thereby augmenting its randomness. The fluctuating degree of freedom guaranteed that the mean and variance of the series changed over time, while the random error introduced noise within the series. The R code for the simulation is presented in Appendix C.

4.2.2. *Resulting Series.* The resulting simulated series $X(t)$ was as illustrated in Figure 3.

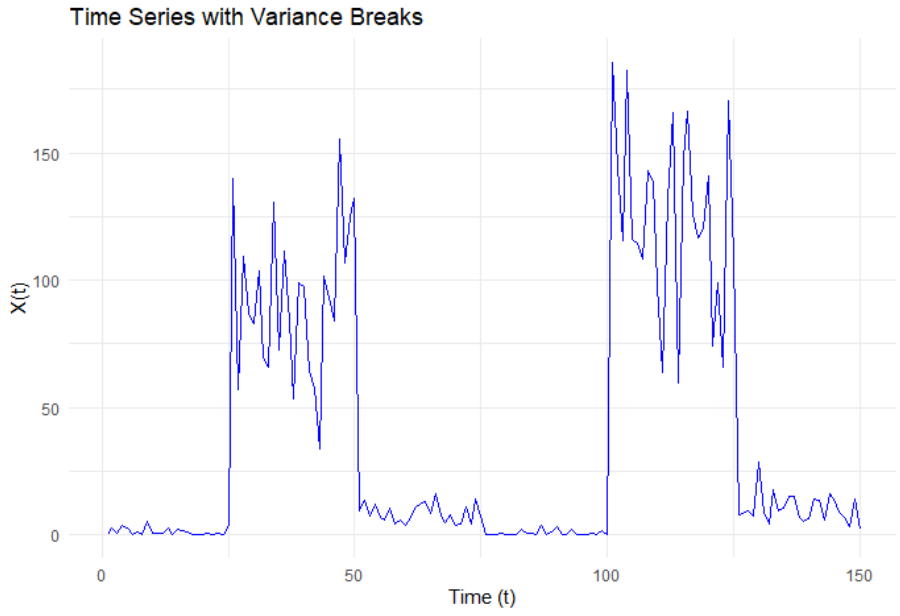


FIGURE 3. Simulated Skewed Time Series

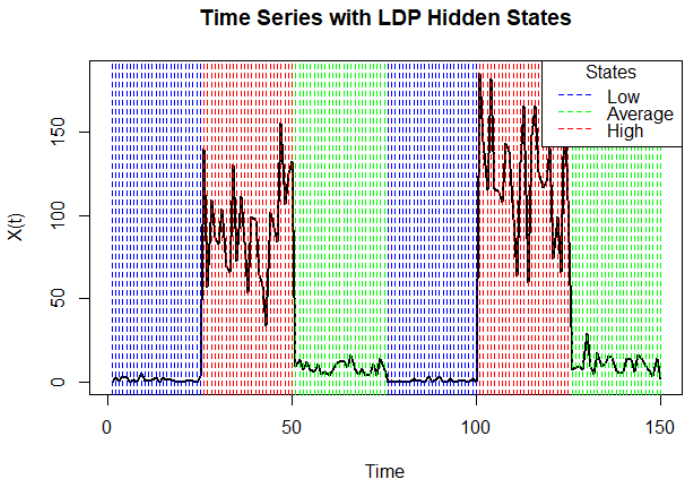
The results in Figure 3 showed a variation in both mean and variance of the series $X(t)$ over time.

4.2.3. *Hidden Markov Models.* We organized the series as a Hidden Markov Model employing both the LDP and EM techniques. In the LDP technique, subsections of size $\delta t = 10$ and a variance of $\sigma^2 = 50$ were employed. The resultant parameter estimates for both LDP-HMM and EM-HMM are depicted in table 2.

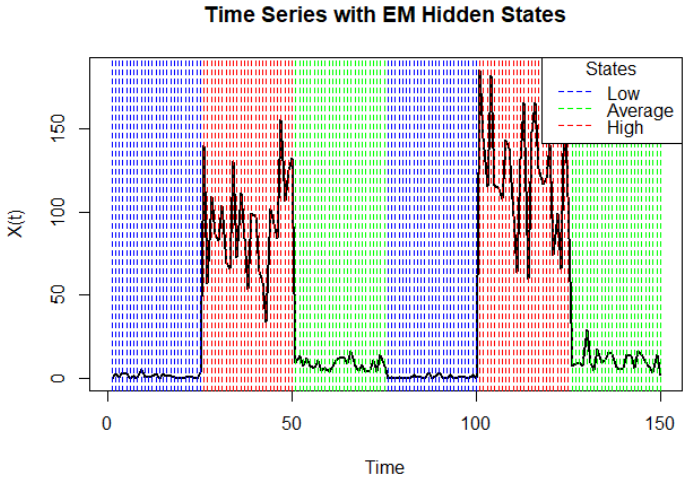
TABLE 2. Parameter Estimates for HMM

Parameter	LDP Estimate	EM Estimates
p_{11}	0.950	0.960
p_{12}	0.000	0.000
p_{21}	0.020	0.024
p_{22}	0.980	0.976
p_{31}	0.000	0.000
p_{33}	0.967	0.960
$\pi_1(1)$	1	1
$\pi_2(1)$	0	0
$\pi_3(1)$	0	0
μ_1	1.033	0.949
σ_1	1.302	1.244
μ_2	8.174	9.408
σ_2	4.809	4.824
μ_3	91.504	108.449
σ_3	50.677	35.945
π_1	0.202	0.273
π_2	0.495	0.454
π_3	0.303	0.273
Log-link	-749.9588	-503.0055
AIC	1527.918	1034.011
BIC	1570.067	1076.16

The results in table 2 showed that the EM algorithm was a better fit the data because it had the smaller AIC and BIC compared to the LDP technique. However, the parameter estimates for the two methods were nearly identical. This evidence suggests that the effectiveness of the LDP technique was almost identical to that of the EM algorithm in modeling the concealed variance regimes of the series. The figures 4a and 4b illustrate a graph depicting the variance regimes of the series derived from fitting the LDP and EM algorithm models utilizing the Viterbi algorithm for decoding in R.



(A) Series with LDP Technique States



(B) Series with EM Algorithm States

The outcomes depicted in Figures 4a and 4b illustrate a comparable classification of variance states by both methods. This evidence suggests that the algorithms were executed in a similar manner.

4.3. Simulated Heteroscedatic Series.

4.3.1. *Simulation Process.* We generated 100 observations for three $Garch(1,1)$ models with varying parameters in R. The initial GARCH model possessed parameters $\mu = 0.3, \omega = 0.1, \alpha_1 = 0.2$, and $\beta_1 = 0.7$. The second GARCH model possessed parameters $\mu = 2, \omega = 10, \alpha_1 = 0.4$, and $\beta_1 = 0.45$. The third GARCH model possessed parameters $\mu = 10, \omega = 80, \alpha_1 = 0.3$, and $\beta_1 = 0.55$. The three series were subsequently amalgamated to create a singular series comprising 300 observations. The simulation codes for the series are located in Appendix D.

4.3.2. *Resulting Series.* The resulting simulated series $X(t)$ was as illustrated in Figure 5.

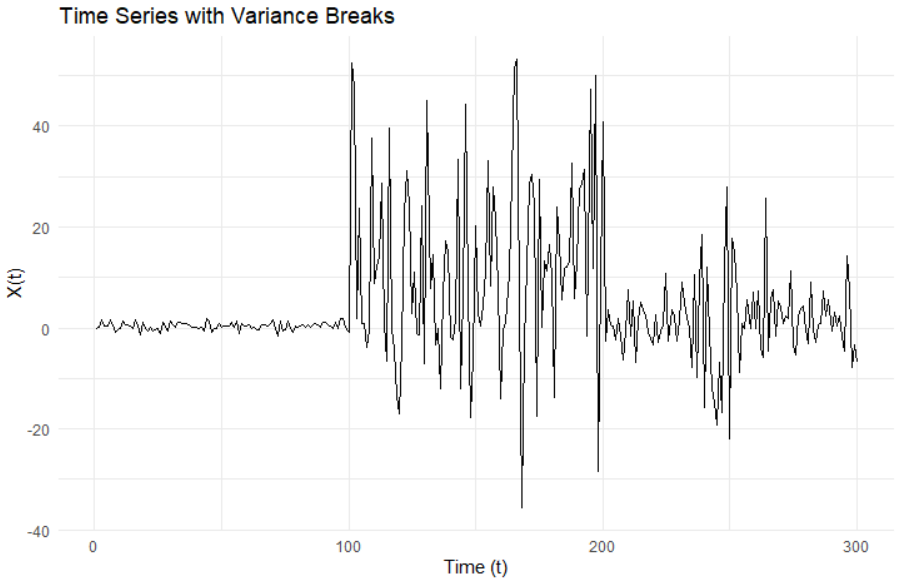


FIGURE 5. Simulated Heteroskedastic Series

The results in Figure 5 showed a variation in variance of the series $X(t)$ over time.

4.3.3. *Hidden Markov Models.* The series was organized as a Hidden Markov Model employing both LDP and EM techniques. The LDP technique employed sub-sections of size $\delta t = 20$

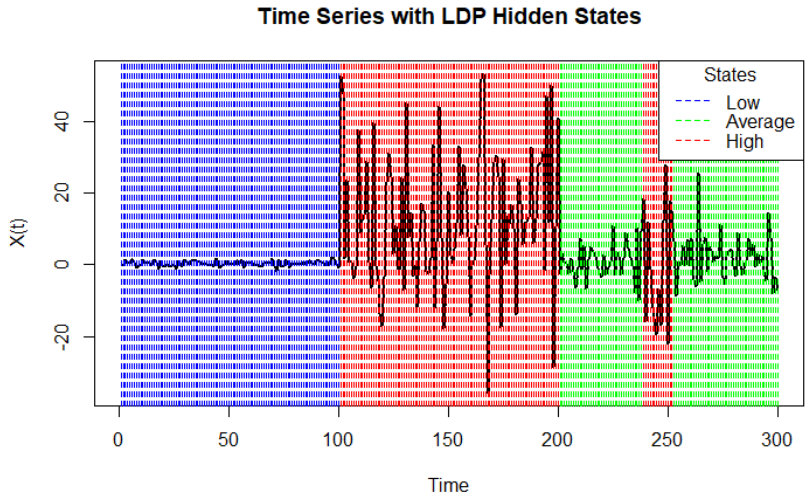
and a contrasting variance of $\sigma^2 = \frac{1}{3}\sigma^2_{300}$. The parameter estimates for both LDP-HMM and EM-HMM are presented in table 2.

TABLE 3. Parameter Estimates for HMM

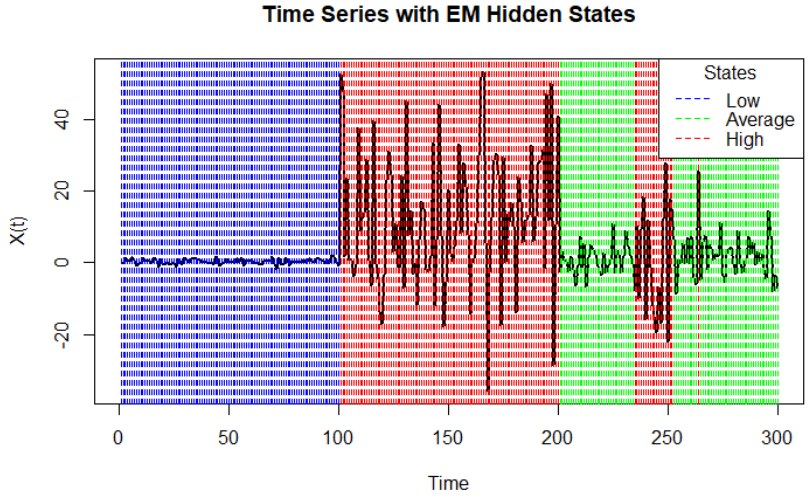
Parameter	LDP Estimate	EM Estimates
p_{11}	0.990	0.990
p_{13}	0.010	0.010
p_{23}	0.013	0.046
p_{22}	0.987	0.954
p_{32}	0.017	0.037
p_{33}	0.983	0.963
$\pi_1(1)$	1	1
$\pi_2(1)$	0	0
$\pi_3(1)$	0	0
μ_1	0.357	0.358
σ_1	0.743	0.739
μ_2	1.463	1.215
σ_2	6.237	4.495
μ_3	10.100	10.303
σ_3	18.177	18.381
π_1	0.000	0.000
π_2	0.568	0.446
π_3	0.432	0.554
Log-link	-1068.301	-890.0506
AIC	2164.603	1808.101
BIC	2216.456	1859.954

The results in table 3 shows that the EM algorithm was a better fit to the data compared to the LDP technique. However, the model parameter estimates for the two methods were nearly

identical. The evidence suggests that the effectiveness of the LDP technique was almost comparable to that of the EM algorithm. Figures 6a and 6b illustrate a graph depicting the variance states of the LDP and EM algorithms, derived from model fitting utilizing the Viterbi decoding algorithm in R.



(A) Series with LDP Technique States



(B) Series with EM Algorithm States

The results in Figures 6a and 6b exhibit nearly identical classifications of variance states by the two methodologies. This indicates that the algorithms exhibited nearly identical performance.

5. Application on Gold Price Data

The majority of nations globally have allocated their wealth in gold. This renders gold essential and among the most valuable commodities globally. The constancy of gold prices indicates economic stability. Nonetheless, fluctuations in gold prices exemplify volatility in the global economy.

The monthly price of gold bars in dollars from 1978 to 2023 was sourced from the Kaggle website, as illustrated in Figure 7.

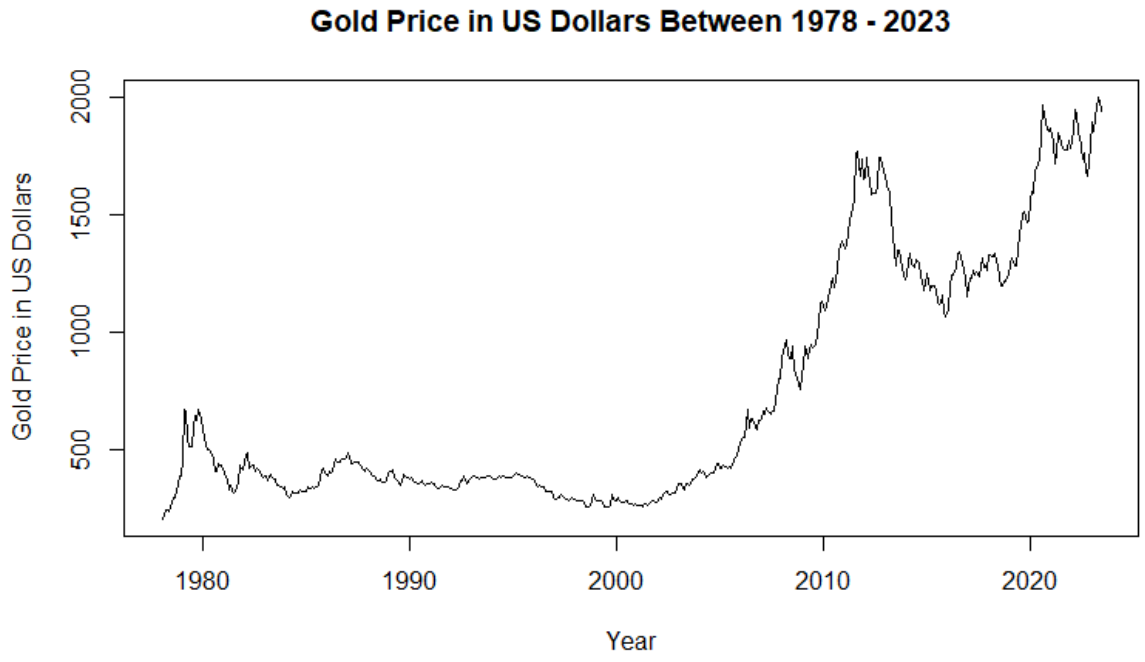


FIGURE 7. Gold Price Series

The data presented in Figure 7 indicates that from 1985 to 2005, fluctuations in the gold price were negligible. The findings also demonstrate that the value of gold has experienced a consistent increase over the years. The value of gold underwent a significant increase between 2007 and 2023. To comprehend the fluctuations in gold prices over the years, the variance in gold values was calculated to ascertain the monthly price changes. The sequence of difference values is depicted in Figure 8.

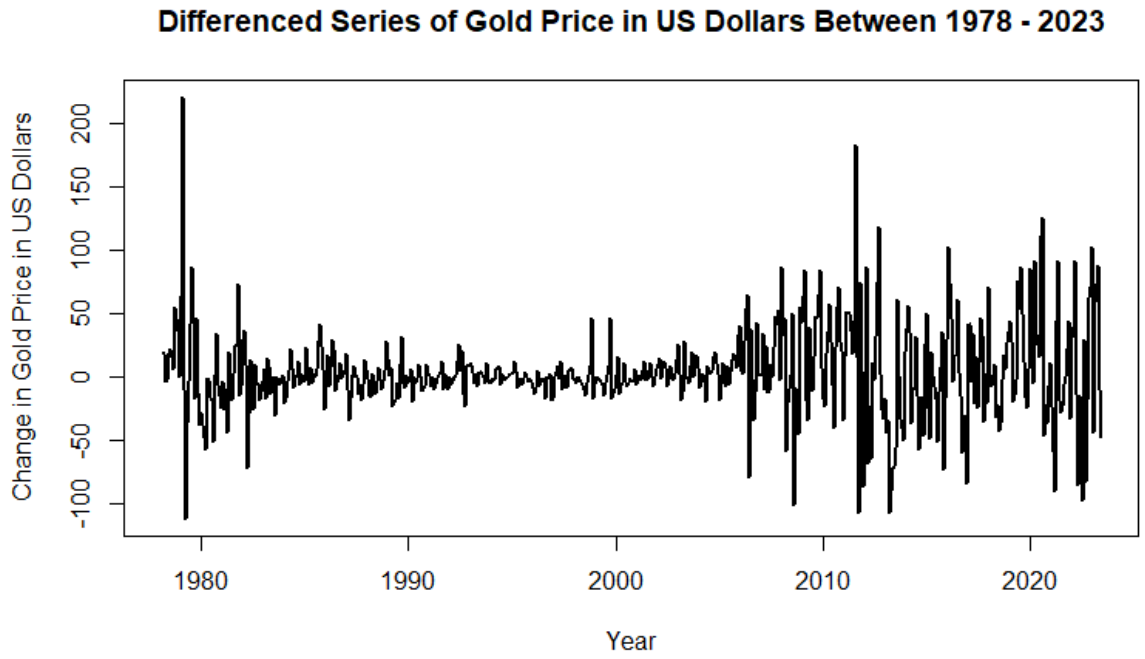


FIGURE 8. Change in Gold Price Series

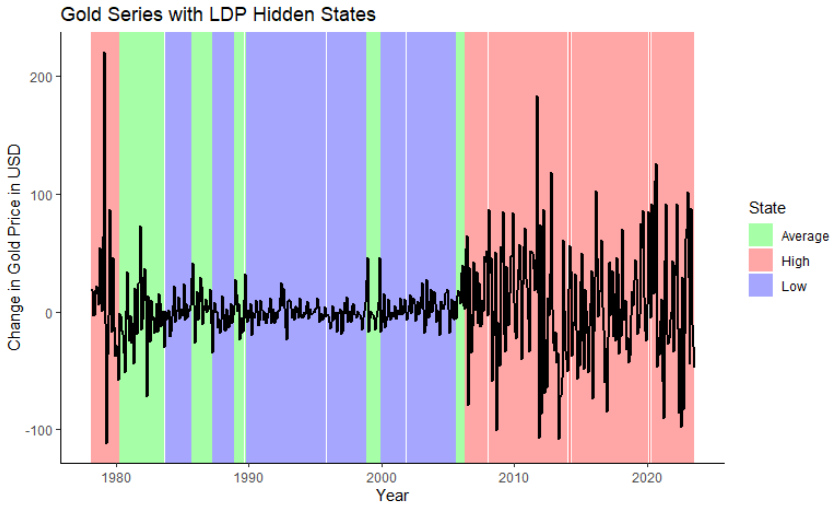
The data presented in Figure 8 indicate that the value of gold exhibited significant volatility prior to 1980 and subsequent to 2007. From 1990 to 2000, the value of gold exhibited reduced volatility, indicating a high degree of stability.

The volatility patterns of gold prices over the years were analyzed by modeling various gold series as a Hidden Markov Model (HMM) utilizing both the LDP technique and the EM algorithm. The LDP method employed a subsection size of two years, utilizing one-third of the total series variance as the benchmark variance. This was executed based on the belief that the overall variance of the series comprised low, medium, and high variance regimes. The parameter estimates of the Hidden Markov Models (HMMs) are presented in Table 4.

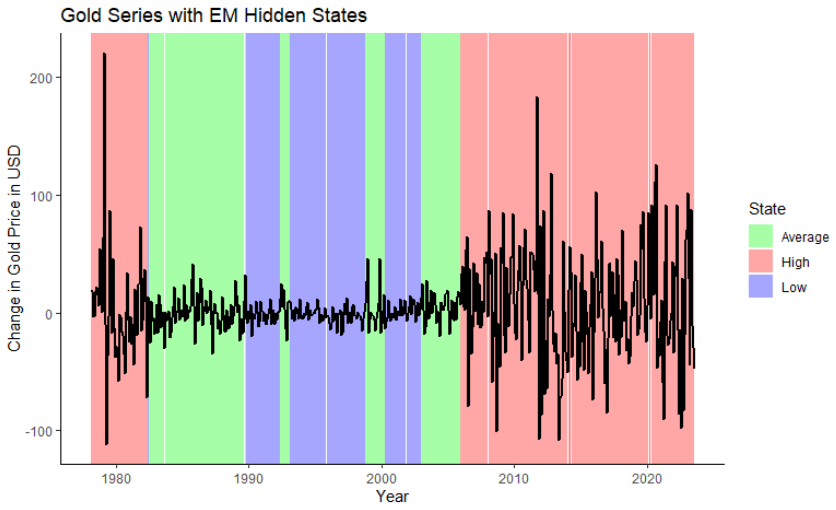
TABLE 4. Parameter Estimates for HMM

Parameter	LDP Estimate	EM Estimates
p_{11}	0.983	0.941
p_{12}	0.017	0.059
p_{21}	0.010	0.054
p_{22}	0.986	0.932
p_{32}	0.009	0.000
p_{33}	0.991	0.995
$\pi_1(1)$	0	0
$\pi_2(1)$	0	0
$\pi_3(1)$	1	1
μ_1	-0.734	-1.365
σ_1	7.913	6.738
μ_2	0.810	2.307
σ_2	24.965	14.573
μ_3	7.657	6.184
σ_3	47.949	48.360
π_1	0.275	0.233
π_2	0.477	0.202
π_3	0.248	0.565
Log-link	-2862.073	-2466.268
AIC	5752.147	4960.536
BIC	5812.358	5020.747

The results in table 4 shows that the EM algorithm was a better fit that the LDP technique because it had the smallest AIC nad BIC. The model parameters were also noticeably distinct from one another. Figures 9a and 9a illustrate a graph depicting the decoded variance regimes derived from the two methods utilizing the Viterbi algorithm for decoding.



(A) Change in Gold Price with LDP Technique States



(B) Change in Gold Price with EM Algorithm States

The results depicted in Figures 9a and 9b indicate that both methods classify high variance regimes identically. A distinction emerged in the categorization of the low and average variance regimes. Moreover, the gold price data indicated that the LDP technique effectively discerned the prevailing volatility regimes. This signifies that the technique can categorize the variance regimes within actual data.

6. Discussion

We employed the LDP technique in simulations and examined its modeling applications. The findings indicated that, similar to the EM algorithm, the LDP technique could delineate the

variance regimes for heteroskedastic time series. Nonetheless, the technique was not a suitable match for the data when compared to the EM algorithm. The discrepancy arises because the LDP technique is intended to identify and delineate the variance regimes within the time series relative to the benchmark variance. The EM algorithm is a method intended to identify the underlying regimes within the series to enhance the observation process. This advantage renders the EM algorithm more suitable, as its aim is to optimize the probability of occurrence in addition to identifying the underlying variance regimes.

The LDP technique exhibits lower computational complexity than the EM algorithm. The EM algorithm employs an iterative procedure that commences with an initial estimation of the series' underlying regimes. This is subsequently succeeded by a series of iterative processes for various underlying states until the optimal set of underlying regimes or states is identified. This complexity requires specialized algorithms and substantial computational power for method implementation. Conversely, the LDP technique disaggregates the series into subsections. The variance of each subsection is estimated to exceed the threshold variance in comparison to the benchmark variance. This can be accomplished manually without requiring specialized software or significant computational resources. This feature renders the LDP technique less computationally intensive than the EM algorithm technique.

The LDP technique is applicable to both small and large datasets for delineating variance regimes in heteroskedastic time series, analogous to the EM algorithm technique. The technique can be utilized across multiple domains, including economics, to identify periods of economic stability (low variance regimes) and economic instability (high variance regimes). In the domain of computer networks, it can be utilized to ascertain the duration of network downtime (low variance regimes) and the duration of elevated network traffic (high variance regimes).

The method can be extended to model non-Gaussian time series, encompassing discrete data. Upon employing the LDP method to identify variance regimes, various emission distributions such as Poisson, t, and generalized extreme value distributions may be utilized to model the emission distribution of the time series. This adaptability renders the methods suitable for various types of time series.

7. Conclusion

In conclusion, the devised LDP technique effectively categorizes the variance regimes of heteroskedastic time series in a manner comparable to the EM algorithm. This suggests that the method may serve as an alternative approach for modeling the variance regimes of multiphase time series models. The method does not utilize an iterative procedure reliant on the initial value, such as the EM algorithm, which may occasionally fail to converge. This advantage renders the method more appropriate for scenarios involving the modeling of variance regimes in heteroskedastic time series when the EM algorithm fails to converge. The method also possesses a defined approach to determine the quantity of underlying variance regimes. This differs from the EM algorithm, which allows the user to select the number of underlying regimes, a choice that can occasionally be erroneous and result in biased estimates. This enhances the method for identifying the number of underlying variance regimes in a heteroscedastic time series and for modeling those regimes. The method enables the modeling of various levels of variance in the multiphased time series and facilitates comparison with the benchmark variance, a capability not afforded by the EM algorithm. This suggests that the model is better equipped to model the variance of heteroskedastic time series than the EM algorithm technique when compared to benchmark variance.

Conflict of Interests

The authors declare that there is no conflict of interests.

References

[1] J.D. Levendis, Structural breaks, in: Time Series Econometrics: Learning Through Replication, Springer, (2023), 175–199.

[2] V. Pata, Fixed point theorems and applications, UNITEXT, Springer, (2019).

[3] M. Deistler, W. Scherrer, ARCH and GARCH models, in: Time Series Models, Springer, (2022), 191–198.

[4] L. Kantar, ARCH models and an application on exchange rate volatility: ARCH and GARCH models, in: Handbook of Research on Emerging Theories, Models, and Applications of Financial Econometrics, Springer, (2021), 287–300.

[5] B.Y. Almansour, M.M. Alshater, A.Y. Almansour, Performance of ARCH and GARCH models in forecasting cryptocurrency market volatility, Ind. Eng. Manage. Syst. 20 (2021), 130–139.

[6] F. Klaassen, Improving GARCH volatility forecasts with regime-switching GARCH, Springer, (2002).

[7] R.F. Engle, T. Bollerslev, Modelling the persistence of conditional variances, Econometric Rev. 5 (1986), 1–50.

[8] C.G. Lamoureux, W.D. Lastrapes, Persistence in variance, structural change, and the GARCH model, J. Bus. Econom. Statist. 8 (1990), 225–234.

[9] S.F. Gray, Modeling the conditional distribution of interest rates as a regime-switching process, J. Financial Econom. 42 (1996), 27–62.

[10] A. Ghezal, I. Zemmouri, et al., Markov-switching threshold stochastic volatility models with regime changes, AIMS Math. 9 (2024), 3895–3910.

[11] L. Maciel, Cryptocurrencies value-at-risk and expected shortfall: Do regime-switching volatility models improve forecasting?, Int. J. Finance Econ. 26 (2021), 4840–4855.

[12] L. Wang, J. Wu, Y. Cao, Y. Hong, Forecasting renewable energy stock volatility using short and long-term Markov switching GARCH-MIDAS models: Either, neither or both?, Energy Econ. 111 (2022), 106056.

[13] C.-Y. Tan, Y.-B. Koh, K.-H. Ng, Dynamic volatility modelling of Bitcoin using time-varying transition probability Markov-switching GARCH model, N. Am. J. Econ. Financ. 56 (2021), 101377.

[14] S.W. Phoong, S.Y. Phoong, S.L. Khek, Systematic literature review with bibliometric analysis on Markov switching model: Methods and applications, SAGE Open 12 (2022), 21582440221093062.

[15] W. Zucchini, I.L. MacDonald, Hidden Markov models for time series: An introduction using R, Chapman and Hall/CRC, (2009).

[16] J.D. Hamilton, Time Series Analysis, Princeton University Press, (2020).

[17] S. Huda, J. Yearwood, R. Togneri, A stochastic version of expectation maximization algorithm for better estimation of hidden Markov model, Pattern Recognit. Lett. 30 (2009), 1301–1309.

[18] S. Huda, J. Yearwood, R. Togneri, Hybrid metaheuristic approaches to the expectation maximization for estimation of the hidden Markov model for signal modeling, IEEE Trans. Cybern. 44 (2014), 1962–1977.

[19] D.W. Stroock, An Introduction to the Theory of Large Deviations, Springer, (2012).

[20] U. Tirnakli, C. Tsallis, N. Ay, Approaching a large deviation theory for complex systems, Nonlinear Dynam. 106 (2021), 2537–2546.

[21] V.M. Gálfi, V. Lucarini, F. Ragone, J. Wouters, Applications of large deviation theory in geophysical fluid dynamics and climate science, Riv. Nuovo Cimento 44 (2021), 291–363.

[22] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Vol. 38, Springer, (2009).

[23] M.A. Arcones, The large deviation principle for stochastic processes: Part I, Theory Probab. Appl. 47 (2003), 567–583.

[24] S.S. Varadhan, Large Deviations, Vol. 27, American Mathematical Society, (2016).

[25] H. Touchette, A basic introduction to large deviations: Theory, applications, simulations, arXiv:1106.4146, (2011).

[26] E.P. Kao, An Introduction to Stochastic Processes, Courier Dover Publications, (2019).

[27] V. Capasso, Introduction to Continuous-Time Stochastic Processes, Springer, (2021).

[28] D.W. Stroock, An Introduction to Markov Processes, Vol. 230, Springer, (2013).

[29] E.B. Dynkin, Theory of Markov Processes, Courier Corporation, (2012).

[30] H. Chen, W. Zhang, C. Yang, H. Cui, Research on marketing prediction model based on Markov prediction, Wireless Commun. Mobile Comput. 2021 (2021), 1–9.

[31] B. Mor, S. Garhwal, A. Kumar, A systematic review of hidden Markov models and their applications, Arch. Comput. Methods Eng. 28 (2021), 1429–1448.

[32] J.-F. Beaumont, L.-P. Rivest, Dealing with outliers in survey data, in: Handbook of Statistics, Vol. 29, Elsevier, (2009), 247–279.

[33] T. Benedict, B. Ramkumari, M. Joseph, O. Fredrick, Classification of variance regime for multiphased time series using large deviation theory, Far East J. Math. Sci. 143 (2025), 85–114.

[34] T. Benedict, B. Ramkumari, M. Joseph, O. Fredrick, Markov switching generalized autoregressive conditional heteroskedasticity (MSGARCH) model based on large deviation principle, Far East J. Theor. Stat. 69 (2025), 295–331.

APPENDIX A. DERIVATION OF RATE FUNCTION

According to Gartner Ellis theorem, if

$$\lim_{\delta t\to\infty} \phi_{\delta t}(\theta) = \phi(\theta)$$

where $\phi_{\delta t}(\theta) = \frac{1}{\delta t}\ln M_v(\theta))$ and $\phi(\theta)$ is a function of θ which is differentiable with respect to θ. Then, LDP holds for $Pr(v \geq a)$ with the rate function $I(a) = \sup(\frac{a\theta}{\delta t} - \phi(\theta))$.
In this case, to determine the rate function, we first find $\phi(\theta)$ and show that it is differentiable with respect to θ.
By definition

$$\begin{aligned}
\phi(\theta) &= \lim_{\delta t\to\infty} \phi_{\delta t}(\theta) \\
&= \lim_{\delta t\to\infty} \frac{1}{\delta t}\ln M_v(\theta)) \\
&= \lim_{\delta t\to\infty} \frac{1}{\delta t}\ln(1-2\theta)^{-\frac{\delta t-1}{2}} \\
&= \lim_{\delta t\to\infty} \frac{1}{\delta t} \times -\frac{\delta t-1}{2}\ln(1-2\theta) \\
&= \lim_{\delta t\to\infty} \frac{\delta t-1}{\delta t} \times -\frac{1}{2}\ln(1-2\theta) \\
&= -\frac{1}{2}\ln(1-2\theta) \text{ if } \theta \in \Re < \frac{1}{2} \wedge \ln(1-2\theta) > 0
\end{aligned}$$

This shows that $\phi(\theta)$ exists. Now we differentiate $\phi(\theta)$ with respect to θ

$$\frac{\partial\phi(\theta)}{\partial\theta} = \frac{1}{1-2\theta}$$

Given that the function is differentiable with respect to θ, we utilise the Legendre-Fenchel Transform to get the rate function $I(a)$. According to Gartner Ellis' theorem, when we apply the Legendre-Fenchel Transform, the rate function is given by $I(a) = \sup(\frac{a\theta}{\delta t} - \phi(\theta))$. Now if we let

$$\begin{aligned}
I(a,\theta) &= \left(\frac{a\theta}{\delta t} - \phi(\theta)\right) \\
&= \left(\frac{a\theta}{\delta t} + \frac{1}{2}\ln(1-2\theta)\right)
\end{aligned}$$

then differentiate $I(a,\theta)$ with respect to θ and equating to 0

$$\begin{aligned}\sup I(a,\theta) &= \frac{\partial I(a,\theta)}{\partial \theta} \\ &= \frac{a}{\delta t} - \frac{1}{1-2\theta}\end{aligned}$$

$$\begin{aligned}\frac{a}{\delta t} - \frac{1}{1-2\theta} &= 0 \\ \frac{a}{\delta t} &= \frac{1}{1-2\theta} \\ 1-2\theta &= \frac{\delta t}{a} \\ 2\theta &= 1 - \frac{\delta t}{a} \\ \theta &= \frac{1}{2} - \frac{\delta t}{2a}\end{aligned}$$

Given that $0 < \theta < \frac{1}{2}$ it follows that $0 < \frac{1}{2} - \frac{\delta t}{2a} < \frac{1}{2}$. This shows that $\frac{\delta t}{2a}$ is a positive number which is less than $\frac{1}{2}$. Then, this means that $\frac{\delta t}{a} < 1$, hence, $a > \delta t$.

Next we substitute for θ in $I(a,\theta)$ to obtain the rate function $I(a)$

$$\begin{aligned}I(a) &= \left(\frac{a\theta}{\delta t} - \phi(\theta)\right) \\ &= \left(\frac{a}{\delta t}(\frac{1}{2} - \frac{\delta t}{2a}) + \frac{1}{2}\ln(1 - 2(\frac{1}{2} - \frac{\delta t}{2a}))\right) \\ &= \left(\frac{a}{2\delta t} - \frac{1}{2} + \frac{1}{2}\ln(1 - 1 + \frac{\delta t}{a})\right) \\ &= \frac{a}{2\delta t} - \frac{1}{2} + \frac{1}{2}\ln\left(\frac{\delta t}{a}\right)\end{aligned}$$

This is the rate function for $Pr(v_{\delta t} \geq a)$. Given that $v_{\delta t} = \frac{(\delta t - 1)V(\delta t)}{\sigma_t^2}$, where δt and σ_t^2 are predetermined constants based on the series $X(t)$.

APPENDIX B. SERIES 1 SIMULATION CODES

```
## Fixing seed for simulation
set.seed(56)
# Load necessary libraries
library(ggplot2)
# Number of observations
n <- 180
#simulation
lambda <- c( rep(75, n / 3),rep(0.6, n / 3), rep(7, n / 3)) *
abs(rnorm(n, mean = 1, sd = 0.3))
ts_data <- rpois(n, lambda)

# Plot the time series
ggplot() +
  geom_line(aes(x = 1:n, y = ts_data), color = "black") +
  labs(title = "Time Series with Variance Breaks",
       x = "Time (t)",
       y = "X(t)") +
  theme_minimal()
```

APPENDIX C. SERIES 2 SIMULATION CODES

```
## Fixing seed for simulation
set.seed(56)
# Load necessary libraries
library(ggplot2)
# Number of observations
n <- 150
#simulation
df <- c(rep(1.2, n / 6), rep(100, n / 6), rep(8, n / 6),rep(0.9, n /
   6),
rep(120, n / 6), rep(12, n / 6)) * abs(rnorm(n, mean = 1, sd = 0.3))
ts_data <- rchisq(n, df)
# Plot the time series
ggplot() +
  geom_line(aes(x = 1:n, y = ts_data), color = "blue") +
  labs(title = "Time Series with Variance Breaks",
       x = "Time (t)",
       y = "X(t)") +
  theme_minimal()
```

APPENDIX D. SERIES 3 SIMULATION CODES

```
library(rugarch)

# Define GARCH(1,1) model specification with user-defined parameters
spec1 <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1,1)),
  mean.model = list(armaOrder = c(0,0)), # No ARMA component
  distribution.model = "norm", # Normal innovations
  fixed.pars = list(mu = 0.3, omega = 0.1, alpha1 = 0.2, beta1 = 0.7)
      # User-defined parameters
)


# Define GARCH(1,1) model specification with user-defined parameters
spec2 <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1,1)),
  mean.model = list(armaOrder = c(0,0)), # No ARMA component
  distribution.model = "norm", # Normal innovations
  fixed.pars = list(mu = 2, omega = 10, alpha1 = 0.4, beta1 = 0.45) #
      User-defined parameters
)


# Define GARCH(1,1) model specification with user-defined parameters
spec3 <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1,1)),
  mean.model = list(armaOrder = c(0,0)), # No ARMA component
  distribution.model = "norm", # Normal innovations
  fixed.pars = list(mu = 10, omega = 80, alpha1 = 0.3, beta1 = 0.55)
     # User-defined parameters
)
```

```r
# Simulate 100 observations
set.seed(123) # For reproducibility
sim1 <- ugarchpath(spec1, n.sim = 100)
sim2 <- ugarchpath(spec2, n.sim = 100)
sim3 <- ugarchpath(spec3, n.sim = 100)

ts_data1 <- as.numeric(sim1@path$seriesSim)
ts_data2 <- as.numeric(sim2@path$seriesSim)
ts_data3 <- as.numeric(sim3@path$seriesSim)

ts_data<-c(ts_data1,ts_data3,ts_data2)
n<-length(ts_data)
# Plot the time series
ggplot() +
  geom_line(aes(x = 1:n, y = ts_data), color = "black") +
  labs(title = "Time Series with Variance Breaks",
       x = "Time (t)",
       y = "X(t)") +
  theme_minimal()
```