Received: 27 April 2017 | Revised: 8 September 2017 | Accepted: 6 November 2017

DOI: 10.1002/hf2.10009

RESEARCH ARTICLE

WILEY HighFrequency

A new analysis of VIX using mixture of regressions: Examination and short-term forecasting for the S & P 500 market

Tatjana Miljkovic[1] | **Indranil SenGupta**[2]

[1]Department of Statistics, Miami University, Oxford, OH, USA

[2]Department of Mathematics, North Dakota State University, Fargo, ND, USA

Correspondence
Indranil SenGupta, Department of Mathematics, North Dakota State University, Fargo, ND, USA.
Email: indranil.sengupta@ndsu.edu

Abstract

A novel approach to the analysis of S & P 500 market fluctuations is proposed using a K-component mixture of regressions model. The Barndorff-Nielsen and Shephard stochastic model is employed where the estimates of jumps of log-returns are governed by Lévy subordinators. Daily VIX and VIX^2 close prices are analyzed as the indicators of log-return volatility and the corresponding variance of the S & P 500 index using the mixture model. The behavior of the S & P 500 market from 1 August 2005 to 31 December 2009 is analyzed and forecasted. A set of rules are provided to predict monthly fluctuation in the S & P 500 market. The procedure used in this paper gives a novel approach for constructing an "indicator"of non-Gaussian jump of an empirical data set in finance using mixture of regression (Gaussian) analysis.

PACS
C40, C53, C65, E17

KEYWORDS
EM algorithm, empirical analysis, Lévy processes, mixture, modeling, VIX price index

1 | INTRODUCTION

Stochastic models are commonly used in assessing price volatility, which is the most important parameter in managing risk, investment portfolios, and asset pricing. Historically, stock market performance was examined periodically from tick-by-tick, half-hour, day, month, quarter, to year for trends. A large volume of research is dedicated to mathematical modeling of the behavior of price volatility with a goal of forecasting the behavior of the market in the next time period. Investors and traders who understand the market's daily, weekly, monthly, quarterly, or annual rhythms may be better equipped to make investment and pricing decisions. In this paper, we focus on the empirical movement of price volatility in the S & P 500 market reflected in financial instruments and derivatives.

The VIX, which is officially known as the Chicago Board Options Exchange (CBOE) Volatility Index, is considered by many to be a gauge of fear and greed in the stock market. More precisely, VIX measures the implied volatility in S & P 500 options. Through the use of a wide variety of option prices, the index gives an estimation of 30-day implied volatility as priced by the S & P 500 index option market. This index can be used to estimate the nature of market movement that the option prices are projecting on the S & P 500 over the next 30-day (or shorter) period. The VIX has an inverse relationship to performance of the S & P 500—a fall of the latter usually corresponds to a rise of the former. After the VIX was introduced, the CBOE introduced futures based on the VIX. Since then, many other financial instruments have been developed or are in the process of development. On 24 February 2006, options on the VIX began on the

CBOE. This gives investors a way to trade the volatility of volatility.

The calculation of VIX depends on the input from all actively quoted S & P 500 index options for the next two standard expirations with reasonable time to expiry (usually 8 days). The elimination of short term expiration takes out some of the "end of contract" fluctuations that may occur in the market. The options from the two expiration series are the at- and out-of-the-money vanilla options. The series of options used extends out-of-the-money as long as there are two successive option strikes that have no bid-ask market posted. The time to expiration is constantly updated. The forward price for the S & P 500 is calculated using put-call parity for the closest at-the-money options. A forward price obtained in this way gives the underlying security price and strike price. This can be used to price the synthetic option, and the implied volatility of such an option is quoted at the VIX. Carr and Madan (1998) and Britten-Jones and Neuberger (2000) show that the VIX can be computed from the prices of call and put options with the same maturity at different strike prices.

The VIX is used as an indicator of the S & P 500 market (Rhoads, 2011). However, it is well known that VIX is much more of a short-term than a long-term market indicator. Spikes in volatility occur quickly, and then the market returns to a stable condition over time. As a result, if the length of time increases, the effectiveness of using VIX decreases as an indicator. Currently, few papers have investigated effectively using VIX to forecast S & P 500 movement. For example, in Lu and Zhu (2010) and Christoffersen, Jacobs, and Mimouni (2010), the nature and importance of volatility risk are discussed by analyzing the pricing of VIX futures. In the paper of Cao and Han (2013), delta-hedging strategies are used related to VIX as a standard risk management technique of option traders in the financial industry. Hardy (2001) developed a regime-switching log-normal model for modeling S & P 500 stock return behavior. The model parameters are derived using maximum likelihood estimation. In this paper, we systematically analyze the VIX by using EM algorithm for mixture of regressions. We show that this analysis is helpful for understanding and predicting the nature of the S & P 500 market.

Mixtures of regression models were introduced in Quandt and Ramsey (1978), as a form of "switching regressions." In that paper, the model parameters of the model are estimated using a technique based on a moment-generating function. In De Veaux (1989), an EM algorithm is considered to fit an application with two regressions. In Jones and McLachlan (1992), mixtures of regressions are studied in a data analysis, and the *Expectation-Maximization* (EM) algorithm is used to fit these models. The problem of determining a number of components in a mixture of linear models is studied by Hawkins, Allen, and Stromberg (2001). In Zhu and Zhang (2004), an asymptotic theory is developed for maximum-likelihood estimators and a hypothesis testing is proposed for mixture of regression models.

This paper presents a novel approach for implementing a mixture of regression technique in analyzing S & P 500 (empirical) data. For the problem studied in this paper, the nature of the world and human being in social contexts plays an important role. Consequently, ontological assumptions are important for this present work. Ontology asserts that social phenomenon is independent from other factors. Specifically, in this paper we propose a completely new approach for analyzing the fluctuations in the S & P 500 market based on the VIX index. This approach is based on a combined knowledge of regression mixtures modeling and stochastic volatility modeling. Inspired by the paper of Maitra (2009) that proposed a staged approach to specifying initial values in partition-optimization algorithms, we develop an efficient EM algorithm for mixture of regressions that uses the stochastic initialization method based on the Euclidean distance. We find that the short-term prediction of the S & P 500 can be obtained by a methodical analysis of this proposed method. In addition to that, this paper gives a completely new empirical approach for constructing an "indicator" of non-Gaussian jump of an empirical data set using mixture of regression (Gaussian) analysis.

The organization of this paper is as follows: Section 2 describes the Barndorff-Nielsen and Shephard stochastic model for modeling stock volatility along with the methodology behind the regression mixture modeling, parameter estimation, a simulation study with initialization of the EM algorithm, and model selection. Data description and the analysis of the results are provided in Section 3. Section 4 provides a brief conclusion.

2 | METHODOLOGY

2.1 | Barndorff-Nielsen and Shephard (BN–S) model for stock and volatility dynamics

In this section, we describe the empirical data under consideration and model the data using Lévy processes. Lévy processes are one of the most efficient building blocks of modern financial models. Popular stochastic processes such as Brownian motions or Poisson processes are special cases of Lévy processes. In our analysis, we consider increasing Lévy processes, which are also called *subordinators*.

The financial time series of different assets share many common features such as heavy tailed distributions of log-returns, aggregational Gaussianity, and quasi long-range dependence. Many such features are successfully captured by models in which stochastic volatility of log-returns is

constructed through an Ornstein–Uhlenbeck (OU) type stationary stochastic process driven by a subordinator, where a subordinator is a Lévy process with no Gaussian component and positive increments. Using Lévy processes as the driving noise, a large family of mean-reverting jump processes with linear dynamics can be constructed. On these processes, various properties such as positiveness or choice of marginal distribution can be imposed. This model was introduced in various works of Barndorff-Nielsen and Shephard (2000, 2001) and is known in recent literature as the BN–S model. Since the pioneering work of Barndorff-Nielsen and Shephard, it is clear that their model has good potential in option pricing and modeling market volatility.

Consider a frictionless financial market where a riskless asset with constant return rate r and a stock are traded up to a fixed horizon date T and $t_0 \le t \le T$. Barndorff-Nielsen and Shephard assumed that the price process of the stock $S = (S_t)_{t\ge 0}$ is defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, P)$ and is given by:

$$S_t = S_0 \exp(X_t), \tag{1}$$

$$dX_t = (\mu + \beta\sigma_t^2)dt + \sigma_t dW_t + \rho dZ_{\lambda t}, \tag{2}$$

$$d\sigma_t^2 = -\lambda\sigma_t^2 dt + dZ_{\lambda t}, \quad \sigma_{t_0}^2 > 0, \tag{3}$$

where the parameters μ, β, $\lambda \in \mathbb{R}$ with $\lambda > 0$ and the stochastic process σ_t govern the volatility of the return. When the parameter $\rho < 0$, a leverage effect is incorporated in the model. Empirically, observations suggest that, for most equities, a fall in price is associated with an increase in volatility. The proposed model is in agreement with this fact. For the model, $W = (W_t)$ is a Brownian motion and the process $Z = (Z_{\lambda t})$ is a subordinator with no deterministic drift. Barndorff-Nielsen and Shephard refer to Z as the *background driving Lévy process* (BDLP). Also W and Z are assumed to be independent and $(\mathcal{F}_t)$ is assumed to be the usual augmentation of the filtration generated by the pair (W, Z). In Equation (2), the Brownian motion and the Lévy process appear as a linear combination, and therefore, the dynamics of the process is linear. Also, the negative sign appearing in Equation (3) makes the associated process mean-reverting. The fact that Equation (3) is driven by Z (instead of W) makes the process non-Gaussian. The unusual timing for $Z_{\lambda t}$ is deliberately chosen so that the marginal distribution of σ_t^2 remains unchanged with respect to values of λ. This gives a flexibility to parameterize separately the distribution of volatility and the dynamic structure (Barndorff-Nielsen & Shephard, 2001).

It is shown in Nicolato and Venardos (2003; Theorem 3.2) that there exists an equivalent martingale measure under which the equations (2) and (3) are transformed into the following equations

$$dX_t = b_t dt + \sigma_t dW_t + \rho dZ_{\lambda t}, \tag{4}$$

$$d\sigma_t^2 = -\lambda\sigma_t^2 dt + dZ_{\lambda t}, \quad \sigma_{t_0}^2 > 0, \tag{5}$$

where

$$b_t = (r - \lambda\kappa(\rho) - \sigma_t^2/2), \tag{6}$$

and W_t and $Z_{\lambda t}$ are Brownian motion and Lévy process, respectively, with respect to the equivalent martingale measure. In the expression for b_t, the cumulant transform for Z_1 under the new measure is denoted as $\kappa(\theta)$ and the risk-free interest rate is given by r. Note that, for a distribution D the cumulant transform is defined as $\kappa(\theta) = \log E[e^{\theta D}]$. For the rest of this section, we assume that the risk-neutral dynamics of the stock price and volatility are given by Equations (1), (4) and (5).

Equation (5) can be written as

$$\exp(\lambda(t - t_0))\left(d\sigma_t^2 + \lambda\sigma_t^2 dt\right) = \exp(\lambda(t - t_0))dZ_{\lambda t},$$

and therefore

$$d(\exp(\lambda(t - t_0))\sigma_t^2) = \exp(\lambda(t - t_0))dZ_{\lambda t}.$$

Thus, the solution of Equation (5) can be explicitly written as

$$\sigma_t^2 = \exp(-\lambda(t - t_0))\sigma_{t_0}^2 + \int_{t_0}^{t} \exp(-\lambda(t - s))dZ_{\lambda s}. \tag{7}$$

Since Z is a subordinator with no drift, σ_t^2 moves up entirely with jumps and then tails off exponentially. Observe that as Z is an increasing process and $\sigma_{t_0}^2 > 0$, the process $\sigma^2 = (\sigma_t^2)$ is strictly positive and is bounded from below by the deterministic function $\sigma_{t_0}^2 \exp(-\lambda t)$. The instantaneous variance of log-returns is given by $(\sigma_t^2 + \rho^2\lambda \mathrm{Var}[Z_1])dt$. Also, it can be proved that the autocorrelation function for the instantaneous volatility is given by $r(h) = \exp(-\lambda|h|)$ (Jongbloed, Van Der Meulen, & Van Der Vaart, 2005).

For the S & P 500, we take VIX as an indicator of log-return volatility. That is, VIX^2 is an indicator of log-return variance. Some work in estimating VIX for future time, based on present VIX, is initiated in Huang and Shaliastovich (2014), Ait-Sahalia, Karaman, and Mancini (2015). The squared VIX index at a given time t is the annualized risk-neutral expectation of the quadratic variation of returns from time t to $t + \tau$, for some $\tau > 0$. For the continuous case, it is given by $\mathrm{VIX}_t^2 = E[\int_t^{t+\tau} \sigma_s^2 ds]/\tau$, where $E(\cdot)$ is the expectation under the risk-neutral measure. We compute below the corresponding expression for the BN–S model.

We assume that there exists a Lévy density associated with the jump measure of Z. Let the random measure associated with the jumps of Z, and Lévy density of Z be given by $J_Z(\cdot,\cdot)$ and ν_Z, respectively. Clearly, the compensator for $J_Z(\lambda dt, dx)$ is $\lambda\nu_Z(dx)dt$. Consequently, for the BN–S model, we can compute the squared VIX index as (Cont &

Tankov, 2004; Issaka & SenGupta, 2017),

$$\begin{aligned} \mathrm{VIX}_t^2 &= E\left[\int_t^{t+\tau} \sigma_s^2 \mathrm{d}s + \int_t^{t+\tau}\int_0^{\infty} \rho^2 z^2 J_Z(\lambda \mathrm{d}t, \mathrm{d}x)\right]/\tau \\ &= E[\int_t^{t+\tau} \sigma_s^2 \mathrm{d}s]/\tau + R, \end{aligned} \tag{8}$$

where

$$R = \rho^2 \lambda \int_0^{\infty} z^2 \nu_Z(\mathrm{d}z), \tag{9}$$

is independent of t. Therefore, with respect to the present model, we obtain

$$\begin{aligned} \mathrm{VIX}_t^2 - R &= E\left[\int_t^{t+\tau} \sigma_s^2 \mathrm{d}s\right]/\tau \\ &= E\left[\int_{s=t}^{t+\tau} \left(\exp(-\lambda(s-t))\sigma_t^2 \right.\right. \\ &\quad \left.\left. + \int_{p=t}^{s} \exp(-\lambda(s-p))\mathrm{d}Z_{\lambda p}\right)\mathrm{d}s\right]/\tau \\ &= (\lambda^{-1}(1-\exp(-\lambda\tau))\sigma_t^2 \\ &\quad + E(Z_1)(\tau - \lambda^{-1}(1-\exp(-\lambda\tau))/\tau. \end{aligned} \tag{10}$$

For the calculations based on empirical data, τ is taken to be 30 days. In view of Equation (10), for the present model we can assume that the relationship between VIX and volatility is linear and it is given by

$$\mathrm{VIX}_t^2 = A\sigma_t^2 + B, \tag{11}$$

where $A = \lambda^{-1}(1-\exp(-\lambda\tau))/\tau$ and $B = E(Z_1)(\tau - \lambda^{-1}(1-\exp(-\lambda\tau))/\tau + R$, where R is given by Equation (9), are constants for a given τ (usually 30 days).

In the data set for a given date, we use VIX for 1 month as the predictor of VIX for the next month. Assuming $T = t_0 + 1$ month $= t_0 + 1/12$ (in the unit of a year), we obtain

$$\sigma_T^2 = \exp(-\lambda/12)\sigma_{t_0}^2 + \int_{t_0}^{T} \exp(-\lambda(T-s))dZ_{\lambda s}.$$

Therefore, using Equations (7) and (11) for $t = t_0$ and $t = T$, we can eliminate $\sigma_{t_0}^2$ and σ_T^2 and obtain the following relation:

$$\begin{aligned} \mathrm{VIX}_T^2 &= \mathrm{VIX}_{t_0}^2 + A(\sigma_T^2 - \sigma_{t_0}^2) \\ &= C_1 \mathrm{VIX}_{t_0}^2 + C_2 + J, \end{aligned} \tag{12}$$

where $C_1 = \exp(-\lambda/12)$ and $C_2 = -B(\exp(-\lambda/12) - 1)$ are constants, and

$$J = A\int_{t_0}^{T} \exp(-\lambda(T-s))dZ_{\lambda s}, \tag{13}$$

measures the fluctuation in prediction of VIX_T^2 with the knowledge of $\mathrm{VIX}_{t_0}^2$. We note that since $C_1 \neq 0$ therefore the "prediction" VIX_T depends on the "predictor" VIX_{t0}.

If there exists a Lévy density $w(x)$ corresponding to the Lévy measure of subordinator Z, the value of J depends on the $w(x)$. In Section 3, based on empirical observations, we propose an "indicator" F for Lévy density of Z. It will be shown that a bigger value of F corresponds to bigger Lévy density which in effect means a more volatile market. Another interesting conclusion derived in Section 3 is that, although the process Z is non-Gaussian, the indicator F, which describes the Lévy density of Z, can be constructed based on Gaussian processes. The indicator F can be used effectively for forecasting the underlying empirical financial data set.

2.2 | A K-component mixture of regression model

Consider the traditional form of the multiple regression model

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \tag{14}$$

where $\boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I})$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters. The matrix, referred to as the design matrix, $\mathbf{X}$ is of size $n \times (p+1)$ and is assumed to have rank equal to $p+1$ (full column rank). The goal of traditional multiple regression is to estimate the parameter vector $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)^\top$. This is accomplished through the least-squares method, which minimizes the sum of squares of deviations for the n observed responses, y_i, from their fitted values, $(\hat{y}_i)$.

Instead of using a single model based on the entire population of data, assume that the population is partitioned into K groups or subpopulations according to the following finite mixture of linear regression model

$$y_i = \begin{cases} \mathbf{x}_i^\top \beta_1 + \epsilon_{i1} & \text{with probability } \pi_1 \\ \mathbf{x}_i^\top \beta_2 + \epsilon_{i2} & \text{with probability } \pi_2 \\ \vdots & \\ \mathbf{x}_i^\top \beta_K + \epsilon_{iK} & \text{with probability } \pi_K, \end{cases}$$

where y_i is the value of the response variable associated with the ith observation, $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ for $i = 1, \ldots, n$ is the transpose of the (p + 1)- dimensional vector of independent variables, β_k for $k = 1, \ldots, K$ is the (p + 1)- dimensional regression coefficient vector of the kth component, and π_k denotes a mixing probability of the kth component satisfying $0 \leq \pi_k \leq 1$, $\forall k = 1, \ldots, K$ and $\sum_{k=1}^{K} \pi_k = 1$. Finally, the random error terms are independent and distributed as $\epsilon_{ik} \sim N(0, \sigma_k^2)$.

For this model, the underlying conditional density of $y_i|\mathbf{x}_i$ is given by

$$f_{\theta(y_i|\mathbf{x}_i)} = \sum_{k=1}^{K} \pi_k \phi\left(y_i|\mathbf{x}_i^\top \beta_k, \sigma_k^2\right). \quad (15)$$

where φ(.) is Gaussian density with mean $\mathbf{x}^\top\beta$ and variance σ^2. The complete parameter vector is denoted as $\theta = (\pi_1, \pi_2, \ldots, \pi_K, \beta_1, \beta_2, \ldots, \beta_K, \sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)^\top$. In addition to estimating θ, we assume that K is not provided and we estimate the number of components in the mixture.

2.3 | Parameter estimation via EM Algorithm

The expectation-maximization (EM) algorithm, proposed by Dempster, Laird and Rubin (1977), is a standard method for finding estimates of the parameters of an underlying distribution in finite mixture models when there are some missing or incomplete data. In our model, missing observations correspond to group identifiers. If we think about the data as consisting of triplets ($\mathbf{x}_i$, y_i, z_i), then z_i is alatent label corresponding to a mixture component from which a pair ($\mathbf{x}_i$, y_i) is drawn.

The EM algorithm is an iterative procedure that consists of two steps: *E*-step (expectation) and *M*-step (maximization).

We start with defining a complete likelihood function given by

$$L_c(\theta) = \prod_{i=1}^{n}\prod_{k=1}^{K} [\pi_k \phi(y_i|\mathbf{x}_i^\top \beta_k, \sigma_k^2)]^{I[z_i=k]} \quad (16)$$

where I[.] is the indicator function with $z_i \in 1, \ldots, K$ and $P(z_i = k \mid \mathbf{x}_i^\top \beta_k, \sigma_k^2) = \pi_k$. The logarithm of $L_c(\theta)$ is known as the complete data log-likelihood given by

$$l_c(\theta) = \sum_{i=1}^{n}\sum_{k=1}^{K} I[z_i = k][\log(\pi_k) + \log(\phi(y_i|\mathbf{x}_i^\top \beta_k, \sigma_k^2))]. \quad (17)$$

The conditional expectation $\mathbf{l_c}(\theta)$ on the current parameter estimates and the observed data is defined as Q-function

$$Q(\theta, \theta^{(s)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(s)}[\log(\pi_k) + \log(\phi(y_i|\mathbf{x}_i^\top \beta_k, \sigma_k^2))] \quad (18)$$

where $\theta^{(s)}$ is the value of θ after the sth EM iteration and

$$\pi_{ik}^{(s)} = \frac{\pi_k{}^{(s-1)}\phi y_i|\mathbf{x}_i^\top \beta_k^{(s-1)}, \sigma^2{}_k^{(s-1)}}{\sum_{k'=1}^{K}\pi_{k'}{}^{(s-1)}\left(y_i|\mathbf{x}_i^\top \beta_{k'}^{(s-1)}, \sigma^2{}_{k'}^{(s-1)}\right)}, \quad i = 1, \ldots, n, k = 1, \ldots, K \quad (19)$$

represents the estimate of the posterior probability on the sth iteration that the ith data point ($\mathbf{x}_i$, y_i) belongs to the kth component of the mixture, computed at the *E*-step.

As a part of the *M*-step, the Q-function is maximized numerically, and parameter estimates are updated to $\theta^{(s)}$. On the sth iteration of the *M*-step, parameter estimates are given by

$$\pi_k^{(s)} = \frac{\sum_{i=1}^{n}\pi_{ik}^{(s)}}{n}, \quad (20)$$

and

$$\beta_k{}^{(s)} = (\mathbf{X}^\top \Pi_k \mathbf{X})^{-1}\mathbf{X}^\top \Pi_k \mathbf{y} \quad (21)$$

where $\mathbf{X}$ is a design matrix, Π_k is a $n \times n$ diagonal matrix of weights (π_{1k}, …, π_{nk}) for $k = 1, \ldots, K$, and y is the response variable vector of size $n \times 1$.

The estimate of variance $\sigma_k^{2(s)}$ on the sth iteration is given by

$$\sigma_k^{2(s)} = \frac{\sum_{i=1}^{n}\pi_{ik}^{(s)}\left(y_i - \mathbf{x_i}^\top \beta_k{}^{(s)}\right)^2}{\sum_{i=1}^{n}\pi_{ik}^{(s)}} \quad (22)$$

The *E*-step and *M*-step are repeated until the relative increase in the log-likelihood function is no bigger than some small pre-specified tolerance value.

2.4 | Initialization of the EM algorithm and simulation

Initialization of the EM algorithm is very important since different starting strategies may lead to different solutions or no solution at all. A large number of initialization procedures are available and discussed in the literature by McLachlan and Peel (2000) and Maitra (2009). Considering the multimodal likelihood function of a finite mixture model and the fact that the EM algorithm is strictly a climbing procedure, the optimal solution is associated with a local maxima. Thus, the choice of initialization is critically important for determining the best global maximizer or the best local maximizer when the likelihood function is unbounded.

In this paper, the EM algorithm is initialized using a stochastic procedure similar to RndEM proposed by Maitra (2009) in clustering data sets. Random k-points are selected from the data, and they are assumed to represent the centers of the k-partitions. Other observations are assigned to these groups based on the minimum Euclidian distance. For defined clusters, the parameter estimates and the log-likelihood are computed. This procedure is repeated m times and the highest value of the log-likelihood function is selected. Finally, a set of parameter estimates are proposed to initialize the EM algorithm based on the highest value of the log-likelihood.

We develop a simulation program in the statistical computing environment R Core Team (2016) and validate the performance of this initialization method in case of a

2-component mixture model with true parameters selected by component and defined as (1) $\beta_1 = (100, 0.1)$ with $\sigma_1 = 30$, and (2) $\beta_2 = (200, 0.2)$ with $\sigma_2 = 10$. Mixing proportions are assumed to be equal for the two groups as $\pi = (0.50, 0.50)$. These parameters are selected to somewhat resemble the S & P VIX-squared price index for the period August 2005 to August 2006. Observed measures are bias computed as $BIAS(\hat{\theta}) = \frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}^{(m)} - \theta)$, and mean-squared error computed as $MSE(\hat{\theta}) = \frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}^{(m)} - \theta)^2$. Table 1 summarizes the results of 1,000 simulation runs.

The results show that this initialization method performs very well leading to an insignificant value of *BIAS* and *MSE*, especially in the case of mixing proportions, which are modeled as predictors of jumps in VIX price volatility.

2.5 | Model selection

In the case of mixture modeling, when the number of components increases, the log-likelihood values as a function of K will monotonically increase. Thus, the models with different number of parameters require criteria that penalize the log-likelihood values by adding a term that increases with the number of parameters, balancing model fit with model complexity. Two popular model selection procedures include the Akaike information criterion (AIC), proposed by Akaike (1974) and the Bayesian information criterion (BIC), proposed by Schwarz (1978). While the AIC considers twice the negative log-likelihood plus a penalty term that is equal to twice the number of free parameters (M) in the model, the BIC approach adjusts the log-likelihood by a penalty term that considers the number of observations (n) in the sample in addition to the number of parameters in the model.

The formulas for AIC and BIC are given by the following equations

$$\mathrm{AIC} = -2l + 2M, \tag{23}$$

$$\mathrm{BIC} = -2l + Mlog(n), \tag{24}$$

where $l(\theta)$ is the maximized log-likelihood, M is the number of free parameters in the model computed as $M = (K-1) + K(p + 1) + K$, and n is the sample size. Our goal is to find the number of components in the mixture that will minimize AIC or BIC. The BIC is often preferred in finite mixture modeling (Fraley & Raftery, 2002).

In addition to AIC and BIC, parametric bootstrapping of the likelihood ratio test statistic (LRTS) can be used to test the null hypothesis $H_0{:}K = K_0$ versus alternative hypothesis $H_a{:}K = K_a$ when $K_a = K_0 + 1$ for some positive integer value K_0 [refer to, 1,000]. Let θ_0 denote vector of parameter estimates under H_0 and θ_1 denote vector of parameter estimates under H_a. Then LRTS is defined as

TABLE 1 Parameter estimates

PAR	*BIAS*	*MSE*
π_1	−0.00010	0.00000
π_2	0.00010	0.00000
β_{11}	−0.04062	0.00165
β_{12}	0.12039	0.01449
β_{21}	−0.00003	0.00000
β_{22}	0.00024	0.00000
σ_1	−0.78932	0.62303
σ_2	−0.18082	0.03269

$$\mathrm{LRTS} = -2\log\lambda = 2[l(\theta_1) - l(\theta_0)]. \tag{25}$$

We expect when λ is sufficiently small or LRTS is sufficiently large that evidence against H_0 will be strong. But in the case of mixture models, LRTS does not have asymptotic null distribution of χ^2 equal to the difference between the number of parameters under H_0 and H_a because the likelihood in the case of mixture models is not sufficiently regular. A large volume of literature is devoted to studying the null distribution of the LRTS for making inferences about the number of components. Some theoretic results are derived in special cases. For instance, McLachlan and Peel (2000) showed that LRTS in the case of testing H_0: $K = 1$ versus $H_a{:}K = 2$ for the K-component univariate normal mixtures has χ^6 rather than χ^4.

In order to overcome the problem of irregularity conditions of LRTS, an effective and easy-to-implement resampling approach is proposed by McLachlan and Peel (2000) that is based on parametric bootstrapping of the LRTS. The bootstrapping algorithm used in the analysis of VIX close prices is outlined in Appendix 1. This algorithm is similar to that implemented by Turner (2000).

We illustrate the bootstrapping method using VIX^2 prices for the period August 2005 to August 2006. Two cases are considered to test the hypothesis about the number of components. Case 1: $H_0{:}K = 1$ versus $H_a{:}K = 2$ and Case 2: $H_0{:}K = 2$ versus $H_a{:}K = 3$. Figure 1 compares the model fit with K = 1 to that model with K = 2.

The histograms of 200 bootstrap test statistics are plotted in Figure 2. It is obvious that for this sample, a 2-component model is the optimal choice. The same model is selected when using the *AIC* or *BIC* approach.

3 | DATA AND ANALYSIS

The complete data set for our analysis is formulated from the observed daily closing VIX data during the period 1 August 2005 through 31 December 2009. For a particular plot, the x-value is taken as the VIX_{t_0} (predictor) for

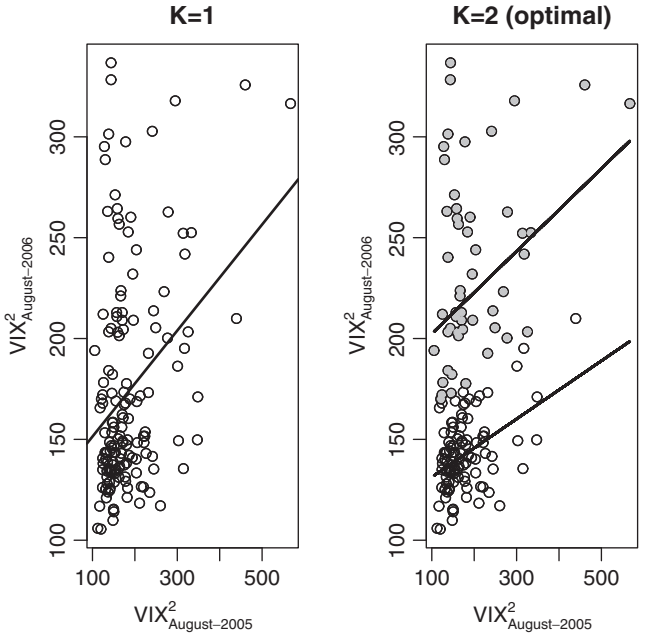

FIGURE 1 1-component vs 2-component model for VIX^2: August 2005–August 2006

14 months and the corresponding y-value is taken as the VIX_T for the next month. Therefore, to construct each plot, we use a data set consisting of 14 months. For example, in the first subset of data, the x-values are VIX_{t_0} during the period 1 August 2005–31 August 2006. Corresponding y-values are VIX_T during the period 1 September 2005 to 30 September 2006. Thus, in order to incorporate the data for a year, observations from 14 months are to be included. For example in the present case, data for the period 1 August 2005–30 September 2006 are necessary. In particular, if the x-value is VIX_{t_0} for 5 August 2005, its corresponding y-value is VIX_T for 5 September 2005. The data for which no pairing is possible with the next month are discarded.

It is clear from Equation (12) that, with respect to the present model, the "jumps" in the log-return (J) are playing a crucial role in determining the relation between VIX_{t_0} and VIX_T. The J term is governed by the Lévy subordinator Z, which appears in the dynamics of the log-return Equation (4). Thus, the fluctuation of the S & P 500 market can be estimated from the analysis of of Equations (12) and (13). Depending on the prominence and intensity of the jump process, the mixture of linear regression model would give various results for the previously described VIX data set. However, as the model of such a data set is given by Equation (12), we can extract important information for the jump process (which are also "jumps" for the S & P 500 market) from the different mixing proportions of mixture of linear regression analysis. We analyze the data set and show that an understanding of the nature of jumps for this model using mixture of linear regression can be used to obtain a premonition of big market fluctuation—such as the S & P 500 crash on 16 September 2008.

3.1 | Classification of S & P 500 market behavior

In this section, we implement the results from mixture of linear regression for the VIX data to analyze the S & P 500 market. We list a set of rules followed by explanations and examples. For the following analysis, K is the optimal number of components for mixture. Let p_{ij}, $1 \le i \le K$, $1 \le j \le i$ be the probabilities of proportions of mixture corresponding to the number of components i (that may or may not be optimal). Also assume that for a given i, if $j' < j''$, then $p_{ij'} \ge p_{ij''}$. We denote the present set of data (14 months) by $\mathcal{P}$ and immediate past set of 14 months data by $\mathcal{P}_p$. Also, for the rest of the paper, we label a 14-month data period by the first month. For example the period 1 August 2005–30 September 2006 will be denoted as "August 2005." Also, in the following when we say "no significant fluctuation in the next month," we mean that the S & P 500 follows more or less the same trend in the following month.

Based on empirical observations, we classify the market fluctuation in the following four cases. We also provide the

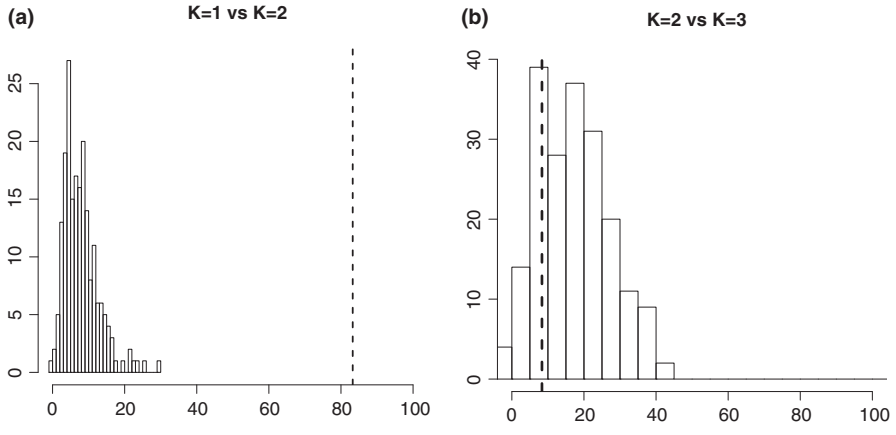

FIGURE 2 Bootstrapping results of the LRTS: left figure shows K = 1 vs K = 2; right figure shows K = 2 vs K = 3. (. . .) represent observed values of the likelihood ratio statistic

expected market behavior for the upcoming month based on empirical observations. Later in this section, we develop a quantitative way of characterizing the following cases.

1. **Case I**. For $\mathcal{P}$, $K = 2$, $p_{21} \leq 0.75$: No significant fluctuation in the S & P 500 during the last 14 months. For the next month, the S & P 500 should not have any big fluctuation. Also, in this case, as p_{21} approaches 0.75, there is more fluctuation at the end of $\mathcal{P}$ and more possibility of big jumps for the S & P 500 during next month.

   Examples of such a case are as follows: August 2005, September 2005, October 2005, November 2005, January 2006, February 2006, March 2006, February 2007, March 2007, April 2007, May 2007. An example of an historical chart (with data from yahoo finance) for the S & P 500 for Case I is provided in Figure 3.

2. **Case II**. For $\mathcal{P}$, $K = 2$, $p_{21} > 0.75$: Significant fluctuation in the S & P 500 during the last 14 months or no fluctuation at all. For the next month, the S & P 500 should have a very big fluctuation. The larger the value of p_{21}, the more the risk of market crash during the next month.

   Examples of such a case are as follows: June 2007, July 2007, August 2007, September 2007. Case II example of historical data for S & P 500 data is provided in Figure 4.

3. **Case III.** For $\mathcal{P}$, $K \geq 3$, $p_{21} \leq 0.70$

- **Case III.A.** If $p_{31} \geq 0.45$: If for $\mathcal{P}_p$, $K = 2$, then big fluctuation in the S & P 500 at the end part of $\mathcal{P}$. Such big fluctuation is expected to continue for the next month. If for $\mathcal{P}$, $K \geq 3$, then significant fluctuation in the S &P 500 at some parts of $\mathcal{P}$. However, for the next month the fluctuations are expected to be stabilized. In both of these cases, the greater the value of K, the more the fluctuation for the S & P 500 during the next month.

  Examples of such a case are as follows: December 2005, April 2006, May 2006, August 2006, September 2006, October 2006, November 2006, February 2008, March 2008, April 2008, May 2008, June 2008, July 2008, September 2008, October 2008, November 2008. An example of an historical chart (with data from yahoo finance) for S & P 500 for Case III. A. is provided in Figure 5.

- **Case III.B.** If $p_{31} < 0.45$: Good fluctuation of S & P 500 at some parts of $\mathcal{P}$. However, for the next month the S & P 500 is expected to be stabilized in terms of fluctuation. The greater the value of K, the more the stabilization of the S & P 500 during the next month.

  Examples of such a case are as follows: December 2006, January 2007, December 2007, January 2008, August 2008. An example of Case III.B. with historical S & P 500 data is provided in Figure 6.

4. **Case IV.** For $\mathcal{P}$, $K \geq 3$, $p_{21} > 0.70$: Some very significant fluctuation in S & P 500 is initiated at the end of 14-month historical data and is expected to continue during the next month. The more the value of K, the more the fluctuation for the S & P 500 during the next month. However, this type of jump is a little less in intensity than the one described in Case II.

   Examples of such a case are as follows: June 2006, July 2006, October 2007, November 2007. An example

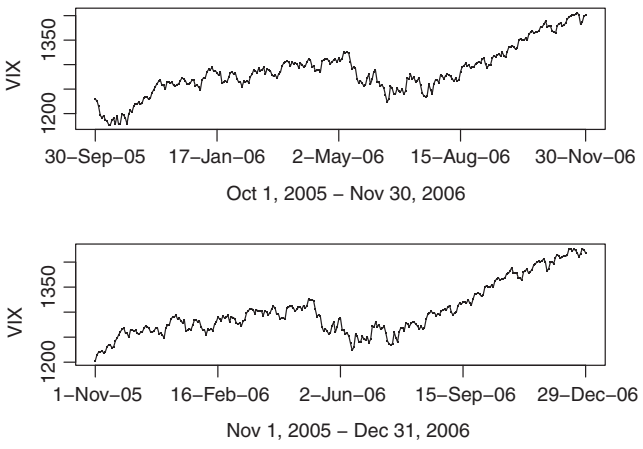

FIGURE 3 October 2005–November 2006 (top), November 2005–December 2006 (bottom)

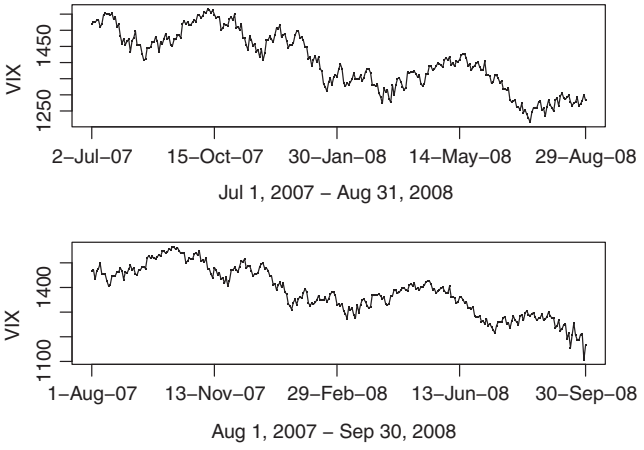

FIGURE 4 July 2007–August 2008 (top), August 2008–September 2008 (bottom)

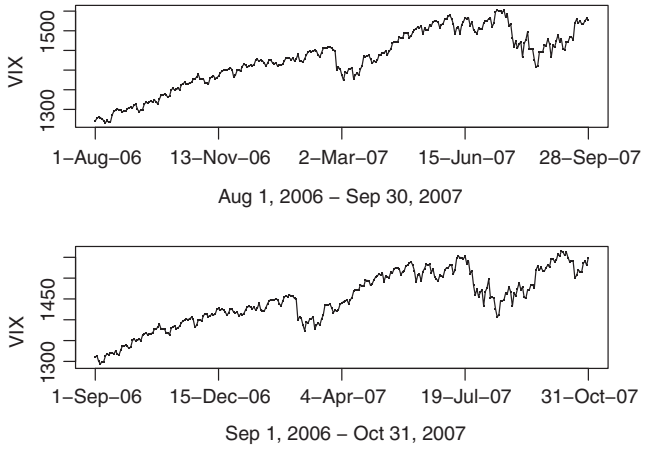

FIGURE 5 August 2006–September 2007 (top), September 2006–October 2007 (bottom)

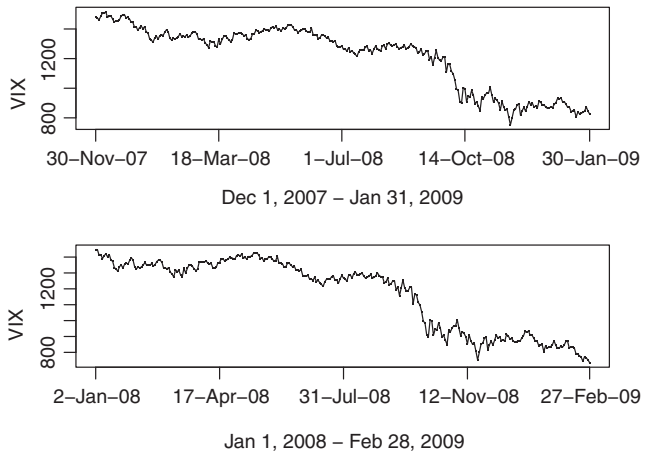

FIGURE 6 December 2007–January 2009 (top), January 2008–February 2009 (bottom)

of historical S & P 500 data for Case IV is provided in Figure 7.

We analyze the above rules heuristically with the help of Equation (12). In the following discussion, "more/less uniformity" means "less/more outliers." It is clear from the nature of the jump process J that it spikes and then descends exponentially. In the presence of reasonably uniform-sized jumps, the splitting of linear mixture of regressions is thus uniform with no "major" mixing proportion. This corresponds to Case I where the S & P 500 index is well behaved. On the other hand, $K \geq 3$ and $p_{31} < 0.45$ imply more optimal number of components for regression model with uniform proportions. Consequently, it signifies higher numbers of possible clustering of jump sizes. However, this does not mean the presence of significant outliers and the S & P 500 is well behaved in this case as well. This situation corresponds to Case III.B. When $K \geq 3$ and $p_{31} \geq 0.45$, this suggests the possibility of some outliers. That means in this case some nonuniform

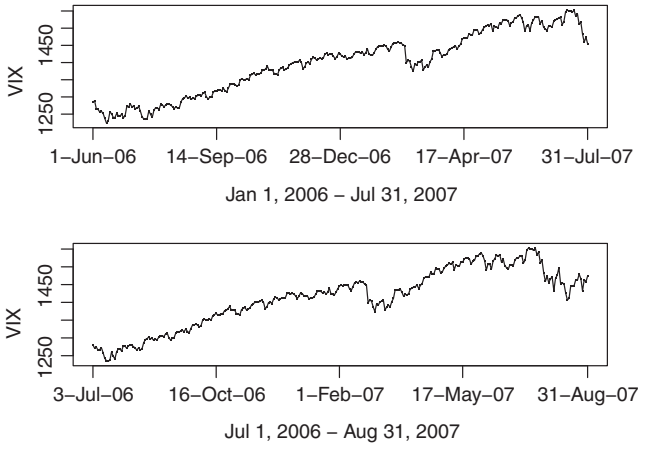

FIGURE 7 June 2006–July 2007 (top), July 2006–August 2007 (bottom)

jump is initiated (or occurred at some point). This corresponds to Case III.A.

The major cases with jumps correspond to Case IV and Case II. For Case IV, $p_{21} > 0.70$, a large number of outliers are present in the S & P 500. However, since $K \geq 3$, such outliers have some uniformity. In other words, the S & P 500 index has nonuniform jumps and that gives rise to the possibility of significant S & P 500 fluctuation for the future month. However, that fluctuation is not as bad as the "crash" for the case when $K = 2$, that is, the situation corresponding to Case II. In Case II, the outliers are significant, but that does not have any good uniformity yet. This situation corresponds to the fact that in the next month this will start building a pattern—in other words, the market will be very volatile. This case may signify a "crash" in the S & P 500 market.

We note that the S & P 500 crash in September 2008 is correctly predicted by the analysis above. The data set July 2007 (which takes into account the data from 07/01/2007 to 08/31/2008) shows Case II behavior, which is a premonition of a crash. Also the start of the housing market crash in August 2007 affected the S & P 500 (though not as badly as September 2008). The data set June 2006 (which takes into account the data from 06/01/2006 to 07/31/2007) shows a Case IV behavior, which correctly captures that situation.

We comment that the estimates of p_{21} or p_{31} described above are based on empirical observations. However, we can provide a heuristic explanation based on Equations (12) and (13). As described in Section 1, the autocorrelation function for the instantaneous volatility at time-lag h is given by $\exp(-\lambda|h|)$. Therefore, from the model it is clear that for sufficiently small values of h the volatility of log-return of S & P 500 prices is correlated. This gives a short-term persistence of volatility and, consequently, for the VIX index.

For Case I, it is clear there are not many outliers. In this case, the jumps (J) are more or less evenly distributed and,

therefore, the S & P 500 is stable. Also, due to the short-term persistence in VIX, it can be concluded that no big fluctuation for the S & P 500 will occur for the next month. Case III.B gives a "splitting effect" for one of the two proportions of Case I. Thus, Case III.B has less uniformity in J than Case I, and consequently, the market is more volatile in the former case. On the other hand, for Case II, there are two possibilities—(i) almost uniformity of jumps, or (ii) a few big outliers are initiated at the end (which means they have no time to form a significant pattern even due to "tail-off" effect of J). In the case (i), since no big jump appeared for the last 14 months, it is highly probable from the model that a big jump is going to occur during the next month. For (ii), due to the persistence of VIX, it is clear that big jumps will occur during the next month. Case IV is the case where there is some "splitting effect" for proportions of Case II. This gives some uniformity in J, which makes it a little more predictable than Case II. Finally, in terms of the structure of J, Case III.A is between Case III.B and Case IV—it has more uniformity of J than Case IV and less uniformity of J than Case III.B. Thus, Case III.A is more predictable compared to Case IV, but less predictable compared to Case III.B. Note that, in all of the above explanation, we use a systematic way of identifying the uniformity/nonuniformity of J and use short-term persistence of volatility to forecast the pattern of J for the upcoming month.

We note that for the BN–S model, jumps of the S & P 500 are governed by the same J (up to some constant) as described before. Therefore, this analysis gives a premonition of big jumps for the upcoming month. The above analysis gives interesting results that cannot be obtained just by looking at the VIX index. For example, the VIX data for the periods 06/01/2006-07/31/2007 or 07/01/2007–08/31/2008 are more or less uniform with no significant sign that may lead to forecast a near future crash. However, that follows from the above analysis.

3.2 | "Indicator" for Lévy density

In this section, based on empirical observation, we propose the following "indicator" for the Lévy density of the governing subordinator Z.

We denote

$$\mathrm{Atan2}(y,x)=\begin{cases}\arctan(\frac{y}{x}), & \text{if } x>0\\ \arctan(\frac{y}{x})+\pi, & \text{if } x<0 \text{ and } y\geq 0\\ \arctan(\frac{y}{x})-\pi, & \text{if } x<0 \text{ and } y<0\\ \frac{\pi}{2}, & \text{if } x=0 \text{ and } y>0\\ -\frac{\pi}{2}, & \text{if } x=0 \text{ and } y<0\\ \text{undefined}, & \text{if } x=0 \text{ and } y=0.\end{cases}$$

We denote K_{opt} to be the optimal number of components for the mixture. Also we assume $K_{opt} \leq N$, for some integer $N > 0$. Let $0<e<0.01$ be a fixed small number. We define the "indicator" as follows:

$$F=\begin{cases}\frac{4p_{21}}{3}+\lfloor\frac{4p_{21}}{3}\rfloor-\epsilon, & K_{\mathrm{opt}}=2,\\ 1+\frac{K_{\mathrm{opt}}}{N}\left[\frac{1}{2}-\frac{1}{2\pi}\mathrm{Atan2}(0.7-p_{21},p_{31}-0.45)\right], & K_{\mathrm{opt}}\geq 3,\end{cases}$$

where $\lfloor x\rfloor$ denotes the floor of x. With appropriate choice of ε, the indicator satisfies: $F < 1$ for Case I; $1 \leq F < 2$ for Case III.A, Case III.B, and Case IV; and $F \geq 2$ for Case II. For any empirical purpose, it is sufficient to take ε = .001 and $N = 4$.

Cases III.A, III.B, and IV can also be distinguished by the values of F. For example, for $K_{opt} = 4$, $F \in [1, 1.25]$ represents Case III.B, $F \in (1.25, 1.5]$ represents Case III.A, and $F \in (1.5, 2)$ represents Case IV. On the other hand, if $K_{opt} = 3$, $F \in [1, \frac{19}{16}]$ represents Case III.B, $F \in (\frac{19}{16}, \frac{11}{8}]$ represents Case III.A, and $F \in (\frac{11}{8}, \frac{7}{4})$ represents Case IV.

Examples of F-values (rounded) for Case I, II, III.A, III.B, and IV are provided in Tables 2, 3, 4, 5, and 6, respectively.

To connect this indicator F with the model described in Section 1, we take examples of well-known subordinators. It is clear from the above analysis that, for forecasting purposes, the Lévy density of Z_t must be an increasing function of the indicator F. It is well known that inverse Gaussian (IG) distributions and variance gamma distributions are self-decomposable, and it is shown in Nicolato and Venardos (2003) that, for BN–S model, if the stationary distribution of σ_t^2 is given by IG (δ, γ) law, then the Lévy density of Z_t is given by $w(x) = \frac{\delta}{2\sqrt{2\pi}} x^{-\frac{3}{2}}(1 + \gamma^2 x)e^{-\frac{1}{2}\gamma^2 x}$, $x > 0$. Thus, $\delta = F$ may be taken and the calibration of data may be obtained with the other parameter γ. This will correspond to jump modeling based on empirical observation. Similarly, if the stationary distribution of σ_t^2 is given by gamma law Γ(ν, α), then the Lévy density of Z_t is given by w

TABLE 2 Examples of F-values (rounded) for Case I

F-value	Time period
0.812	August 2005
0.812	September 2005
0.932	October 2005
0.946	November 2005
0.906	January 2006
0.852	February 2006
0.826	March 2006
0.972	February 2007
0.746	March 2007
0.759	April 2007
0.906	May 2007

TABLE 3 Examples of *F*-values (rounded) for Case II

F-value	Time period
2.039	June 2007
2.331	July 2007
2.199	August 2007
2.146	September 2007

TABLE 4 Examples of *F*-values (rounded) for Case III.A

F-value	Time period
1.243	December 2005
1.211	April 2006
1.292	May 2006
1.426	August 2006
1.400	September 2006
1.228	October 2006
1.196	November 2006
1.449	February 2008
1.243	March 2008
1.383	April 2008
1.336	May 2008
1.366	June 2008
1.275	July 2008
1.349	September 2008
1.330	October 2008
1.489	November 2008

TABLE 5 Examples of *F*-values (rounded) for Case III.B

F-value	Time period
1.153	December 2006
1.151	January 2007
1.024	December 2007
1.153	January 2008
1.160	August 2008

TABLE 6 Examples of *F*-values (rounded) for Case IV

F-value	Time period
1.602	June 2006
1.563	July 2006
1.415	October 2007
1.667	November 2007

$(x) = \nu\alpha e^{-\alpha x}$, $x > 0$. Thus, $\nu = F$ may be taken, and the calibration of data may be obtained with the other parameter α. This will also correspond to jump modeling based on empirical observation. This procedure gives a novel approach for constructing an indicator (F) of non-Gaussian jump (J) of an empirical data set using mixture of regression (Gaussian) analysis.

4 | CONCLUSION

This paper incorporates knowledge of mixture of regression modeling for the VIX index in analyzing and forecasting jumps in S & P 500 prices modeled by the BN–S stochastic model. Four cases of jumps are examined as predictors of the S & P 500 fluctuations for the short-term forecast during the period 1 August 2005 to 31 December 2009, using 14-month observed period data. All of the cases are based on the mixing proportions estimated for the optimal K-component linear regression model. We find that the way VIX prices cluster can tell us about the fluctuations and behavior of the S & P 500 market. We also find that the VIX^2 prices can rarely be described using a single-component regression model as they show tendency of clustering.

When we observe more uniform partitions in the 2-component mixture model for VIX prices, jumps are more uniform and they are expected to continue in the next month. In the case of a bigger gap between partitions in the 2-component model, there is an indication of greater fluctuations toward the end of the observed period, which leads to a higher probability of having big jumps in the next month. This situation was confirmed for the period June 2007–September 2007 when a big market crash occurred. In cases where more than a 2-component regression model is found optimal, we set the rules for analyzing VIX data and predicting jumps in market fluctuations.

Our results have implications in calculations of short-time horizon Value-at-Risk (VaR) and Tail-Value-at-Risk (TVaR). The VaR is known as a "key risk indicator" and represents a quantile of the estimated profit-and-loss (P & L) distribution for a portfolio of financial assets. The VaR with a given probability level provides information to actuaries about "the chance of an adverse outcome" for a particular financial asset. On the other hand, TVaR provides the average excess loss under adverse outcomes in cases when the VaR quantile is exceeded. These adverse outcomes are referred to as "bad times" and are often linked to the performance of the stock market. Both VaR and TVaR are considered forward-looking over a fixed short time period (e.g., several days or a month) with a high probability. If a portfolio contains holdings of the S & P 500 index, then VIX could be used directly in the calculation of VaR and TVaR. The findings of this research could be used in forecasting short-term jumps in VIX prices, resulting in more accurate risk evaluation of the underlying

investment portfolio, as well as calculations of VaR and TVaR measures.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their careful reading of the manuscript and for suggesting points to improve the quality of the paper.

ORCID

Indranil SenGupta http://orcid.org/0000-0002-4842-1352

REFERENCES

Ait-Sahalia, Y., Karaman, M., & Mancini, L. (2015). The term structure of variance swaps and risk premia. (Working Paper).

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Barndorff-Nielsen, O., & Shephard, N. (2000). Modelling by Lévy processes for financial econometrics. Theory and Applications (No. 2000-W3).

Barndorff-Nielsen, O., & Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of Royal Statical Society: Series B (Statistical Methodology)*, *63*(2), 167–241.

Britten-Jones, M., & Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, *55*(2), 839–866.

Cao, J., & Han, B. (2013). Cross-section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, *108* (1), 1–10.

Carr, P., & Madan, D. (1998). Towards a theory of volatility trading. *Volatility: New estimation techniques for pricing derivatives*, *29*, 417–427.

Christoffersen, P., Jacobs, K., & Mimouni, K. (2010). Volatility dynamics for the s & p 500: Evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies*, *23*(8), 3141–3189.

Cont, R., & Tankov, P. (2004). *Financial modelling with jump processes*. Boca Raton, FL: Chapman & Hall.

De Veaux, R. (1989). Mixtures of linear regressions. *Computational Statistics & Data Analysis*, *8*, 227–245.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society B*, *39*, 1–38.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, *97*(458), 611–631.

Hardy, M. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal*, *5*(2), 41–53.

Hawkins, D. S., Allen, D. M., & Stromberg, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis*, *38*(1), 15–48.

Huang, D., & Shaliastovich, I. (2014). Volatility-of-volatility risk. Working paper, University of Pennsylvania.

Issaka, A., & SenGupta, I. (2017). Analysis of variance based instruments for Ornstein-Uhlenbeck type models: Swap and price index. *Annals of Finance*, *13*(4), 401–434.

Jones, P., & McLachlan, G. (1992). Fitting finite mixture models in a regression context. *Australian & New Zealand Journal of Statistics*, *34*(2), 233–240.

Jongbloed, G., Van Der Meulen, F. H., & Van Der Vaart, A. W. (2005). Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, *11*(5), 759–791.

Lu, Z., & Zhu, Y. (2010). Volatility components: The term structure dynamics of vix futures. *Journal of Futures Markets*, *30*(3), 230–256.

Maitra, R. (2009). Initializing partition-optimization algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *6*, 144–157.

McLachlan, S., & Peel, D. (2000). *Finite Mixture Models*. Hobuken, NJ: John Wiley & Sons.

Nicolato, E., & Venardos, E. (2003). Option pricing in stochastic volatility models of the ornsteinuhlenbeck type. *Mathematical Finance*, *13*(4), 445–466.

Quandt, R., & Ramsey, J. (1978). Estimating mixtures of normal distributions and switching regression. *Journal of American Statistical Association*, *73*(364), 730–738.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/

Rhoads, R. (2011). *Trading VIX derivatives: Trading and hedging strategies using VIX futures, options, and exchange traded notes*. Hoboken, NJ: John Wiley & Sons.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Turner, T. (2000). Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(3), 371–384.

Zhu, H., & Zhang, H. (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(1), 3–16.

How to cite this article: Miljkovic T, SenGupta I. A new analysis of VIX using mixture of regressions: Examination and short-term forecasting for the S & P 500 market. *High Freq*. 2018;00:1–13. https://doi.org/10.1002/hf2.10009

APPENDIX

ALGORITHMS

EM Algorithm

m←1

l_{old}←−Inf

{REPEAT Loop} M times

Select K-points at random $(\mathbf{x}_j, y_j)$, $j = 1,\ldots,K$ from n points.

Assume they represent centers of K groups.

Assign observations to these groups based on the smallest Euclidian distance to a cluster center

Estimate $\theta_m^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, \ldots, \pi_K^{(0)}, \beta_0^{(0)}, \beta_1^{(0)}, \ldots, \beta_K^{(0)}, \sigma_1^{2(0)}, \sigma_2^{2(0)}, \ldots, \sigma_K^{2(0)})^\top$

Calculate $l(\theta_l^{(0)})$

$\theta_{opt} \leftarrow \arg\max_{m=1,\ldots,M} l(\theta_m^{(0)})$

$l_{new} \leftarrow l(\theta_{opt})$

m←m+1

end {REPEAT Loop}

s←1

WHILE$(|(l_{new} - l_{old})/l_{new}| \geq eps)$ DO

l_{old}←l_{new}

{E-step}

Compute posterior probabilities $\pi_{ik}^{(s)}$for $i = 1,\ldots, n$ and $k = 1, \ldots K$.

{M-Step}

Update $\theta^{(s)} = (\pi_k^{(s)}, \beta_k^{(s)}, \sigma_k^{2(s)})$

(Continues)

(Continued)

EM Algorithm

Calculate $l(\theta^{(s)})$

s←s + 1

end WHILE

return θ and l_{new}

Algorithm for Bootstrapping LRTS

l←1

{REPEAT loop}

For $K = K_0$ under H_0 compute $\log L(\hat{\theta}_0)$.

For $K = K_a$ under H_a compute $\log L(\hat{\theta}_1)$.

Calculate $Q = -2(\log L(\hat{\theta}_0) - \log L(\hat{\theta}_1))$. This is likelihood ratio test statistic (LRTS).

l←l + 1

end {REPEAT Loop}

{FOR loop} N times

i←1

simulate data from $K = K_0$ model

using the simulated data fit K-and $(K + 1)$ component models

compute bootstrap LRTS, denoted as $Q_{boot}(i) = -2(\log L(\hat{\theta}_K) - \log L(\hat{\theta}_{K+1}))$

i←i+1

end {FOR Loop}

compute the p-value as $p = \frac{1}{(N-1)} \sum_{i=1}^{N} I(Q \leq Q_{boot}(i))$; $I(.)$is an indicator function

return the p-value, Q, Q_{boot}.