# Analysis of Autoregressive hidden Markov model under asymmetric Laplace distribution

Li-Yin Young, Yu-Jui Huang

*Department of Applied Mathematics, University of Colorado Boulder*

**Abstract**

In this paper, we incorporated the observed features of human behavior in stock market into model. The trading volume is taken as the signal of overconfidence in our model for initialized the probability of state and its transition. We employ the autoregressive model with hidden Markov trading strategies, because Hidden Markov model(HMM) in modeling dynamic systems is used extensively for pattern recognition and classification problems in a given time series. The forecasting of stock price is via discrete-valued Markov process under an asymmetric Laplace distribution for forecasting the stock prices. The empirical evidence of our model showed that our model is quite effective and practical in stock market forecasting.

*Keywords:* hidden Markov Model, Autoregressive model, financial time series, stock market forecasting, behavioral finance

## 1. Introduction

The ongoing explosive growth of computing performance and reliability has led to recent advances in machine learning that have yielded transformative results across diverse scientific disciplines. However, data forecasting that involves a pattern that typically alternates with random fluctuations has been largely been beyond the capability of traditional data-driven models. Numerous research studies have applied artificial neural networks(ANN) to find the optimal prediction model for stock prices [[1], [2], [3]]], with these ANNs trained in a batch learning setting, which requires the entire training data set [4]. However, the very nature of the stock market dictates that any analysis tool should be capable of dynamically adapting to temporal patterns in data [ [5], [6]]. Therefore, the batch learning is not suitable for forecasting stock price.

HMM is a tool for determining the hidden state by obtaining the probability distribution of the observing signal [7] which had great success in temporal pattern recognition and unsupervised learning [8]. These aspects have been applied for tracing the volatility of the temporal pattern in stock market[[5], [9], [10]], but they did not develop the estimation algorithm presented in the paper.

Parameter initialization is an important issue in training HMM. Improperly initialized models leads to poor alignment between labels and data in the training phase resulting in the bad estimates of parameters of the HMM. Odean's [11] provided the mathematical model to present that the overconfidence lead to the increase of trading volume and volatility of stock price because it affects people how to value their information. In other words, overconfidence cause people overweight the information. Although overconfidence is only one of the many ways by which disagreement among investors may arise, many previous research [[12], [13], [14], [15], [16], [17], [18]] suggested by some experimental studies of human behavior and generates a mathematical framework that is relatively simple. Therefore, we decide to model the overconfidence as the signal for Hidden Markov model(HMM) based time series model in initialized phase. Our model explores the hidden state by Hidden Markov model(HMM) which observes the change in the trading volume based on the aforementioned studies and the results in Figure 1. The performance of individual stock is said to be in state 0 if the stock price is rising and in state 2 if the stock price is falling with the transition between these two states models as a outcome of a second-order Markov process. Once we initialized the probability of state, ARHMM switch to observe stock return as a signal for hidden state.

To make the model furthermore be capable to estimate the stock price, we decided to introduce regime switching autoregressive models. In HMM, it assumes that the sequence of hidden states satisfies the Markov property which means that the current state is independent of all the states prior to $t-1$. Therefore, it is suitable for modeling first-order hidden Markov autoregressive model (ARHMM(1)). The basic approach is to use Hamiltian 's [19] method to calculate the smoothed probabilities for adjusting the parameters in an autoregressive model. Applied autoregressive hidden Markov model (ARHMM) had been applied to predict the S&P 500 index [[5], [20]] . Caccia et al.[20] compared the actual price with the theoretical prices which is obtained from the ARHMM to evaluate whether the stock index is overpriced or underpriced.

The other important contribution of the papers is that our model is observe the past stock return based on AL distribution. While Gaussian distribution has dominated the field of empirical finance for a long time, the empirical distribution of price change are usually asymmetric and steep peak at the origin to be relative to samples from Gaussian distribution[ [21], [22]] . The skewed generalization of the Laplace distribution and its application has been studied by various authors [[23], [24], [25], Jayakumar[26], [27], Trindade[28], [29]]. Further,our empirical results validate these aforementioned results. We can observe from Figure 1 that the empirical distribution of the stock returns themselves is not normal distribution. Instead, it is typically quite fattailed and more peaked around the center[26] as shown in Figure 1. Although there is considerable literature for the theory and application of AL distribution, the application for modeling financial time-series data are still rather limited. We used the approaches provided in [[26], [27]] to obtain the parameters for the ARHMM model with AL distribution. Trindade et al. presented the case study in [27] that AL noise models tend to have a grader degree of stability for skewed distribution in terms of numerical view.

The goal in our paper is to propose a new algorithm considering the behavior of human and price for forecasting stock prices. The new algorithm adjusts the autoregressive parameters based on the outcome of a discrete-valued Markov process of the trade volume under an asymmetric Laplace distribution. The paper is organized as follows. In Section 2, it discussed why we should expect to find overconfidence in stock market. Section 3 described the structure of the model and showed how to implement the EM algorithm for the parameter optimization. In Section 4, we report the results of the experiments performing on computer-generated data. Finally, Section 5 is devoted to the summary and conclusions.

## 2. Initialization: The relation between trading volume and stock price changes

### *2.1. Data*

The data we consider range from May 29, 2018 to May 24, 2019; a total of 249 daily values. The data set comprises daily market price and trading volume series for five stocks: FB, AAPL, GOOGL, AMZN and AMJ. We choose these markets because they are well established and large for our statistical tests. We calibrate the dataset to every 20 days for each date's training data set. For example, we trained an ARHMM using the daily stock data of Amazon for the period from 1 June 2018 to 20 June 2018 to predict the closing price on 21 June 2018.

Figure 1 presents the distribution of each one of the stock return in our experiment. The result shows that the distribution of returns have fat tail compared with normal distribution. It implies that there exist greater outlier potential in the distribution of stock return.

### 2.2. Trading volume and price change

In recent studies, researcher have found the positive relation between trading volume and the absolute price change by statistics [[30], [16],[31], [32], [33]] . Chen et al. [31] used bivariate autoregressions to test for causality between the two variables trading volume and stock returns

$$V_t = \alpha_0 + \sum_{i=1}^{5} \alpha_i V_{t-1} + \sum_{j=1}^{5} \beta_j R_{t-j} \tag{1}$$

$$R_t = \alpha_0 + \sum_{i=1}^{5} \gamma_i R_{t-1} + \sum_{j=1}^{5} \delta_j V_{t-j} \tag{2}$$

where $V_t$ is the detrended trading volume at time $t$ and $R_t$ is the detrended stock returns at time $t$. The empirical result shows both the coefficients $\beta$ and $\delta$ for both the linear and nonlinear time trends in trading volume series are statistically significant and the model fit is high. Therefore, there is a feedback relation between returns and trading volume.

Pathirawasm [33] further find that monthly stock returns are positively related to the change in trading volume. In Pathirawasm [33] experiment, he grouped the stock return as high trading volume, middle trading volume and low trading volume and implement the t-test to examine the significance between stock and trading volume.

Previous studies provide the evidence that there is a contemporaneous positive relation relation between trading volume and returns. Therefore, we decided to use volume to initialize the probability of state.

The proposed initializing approach contains three steps. At first, the signals are split into different regimes based on different behaviors. Second, similar regimes of the signal are grouped together by K-mean clustering method. Thirdly, the achieved labeled regimes are assume to be the hidden state. Finally, the parameters are estimated by calculating statistical occurrences of the observed signal and the estimated hidden states, then used as initial input of the conventional Baum-Welch algorithm.

We aim to find the feature that captures the regime switching in stock market.

In our model, the individual stock is said to be in state 0 if the stock price is rising and in state 1 if the stock price is falling with the transition between these two states models as a outcome of a second-order Markov process. The state process is the homogeneous Markov chain on $t \in \{1, \ldots, 20\}$ with transition matrix $a_{i,j}$. We aim to find the most likely state at time t.

$$\pi_t = P(s_t = i | y_1, \ldots, y_{t-1})$$

The ARHMM model is based on the linear structure of two states which refer to price increasing or decreasing. The trading volume of the stock price is extracted and clustered by K-means clustering. A cluster refers to a collection of data points aggregated together due to certain similarities. K-means clustering partitions $n$ the daily trading volume into $k$ clusters in which each observation belongs to the cluster with the nearest mean of each group in each calibration. The cluster of each observation serves as its beginning state. Our model assume the group come from same regime leads to have the same state.

## 3. ARHMM as a predictor

In this section we considered ARHMM models for forecasting the next day stock price. Since our goal is maximize the likelihood function for hidden state with respect to the stock return parameters, the observed signal is stock return in training state. The emission probability is concerned in AL distribution because of the empirical evidence of the distribution of individual stock return. According to Figure 1, it showed that the empirical distribution of the stock returns themselves is not normal distribution We first derived the framework of probability distribution for HMM and express the PDF for stock price. Then, we proposed a parameter estimation approach. Finally, we will discuss the prediction problem.

### *3.1. Notation*

For the rest of the paper, the following notation will be used regarding ARHMM.

- State: $s_t \in \mathbb{R}, N \in \mathbb{N}, s_t \in \{1, \ldots, N\}$;
- Transition matrix $a_{i,j} \in \mathbb{R}^{T \times i}$ where $i$ and $j$ are the state.
- Mean value of the stock price: $\mu^{(i)}$ in state $s$;
- Covariance matrix $\Sigma \in \mathbb{R}^{T \times N}$
- Autoregressive lag: $p \in \mathbb{R}^N$.

### *3.2. A probabilistic framework of HMM*

Let y denote a vector of the observations and $T$ denote a finite time interval $[0, T]$, where $T > 0$.

$$y = \{y_1, \ldots, y_T\}$$

We assume that there exist the occasional shift in the level, variance or autoregressive dynamics of $y$. Suppose that there are a total of K possible states from which the observations $(x_1, \ldots, x_T)$ have been drawn. To model this concept, the unobserved scalar variable which are unable to be directly estimated via observed data. is referred as the "state" of the process. Jurafsky et al. [34] described the HMM as the augmenting Markov chain. Markov chain is the probabilities of the sequence of states. If the state is currently in $i$ state, then it moves to $j$ state with probability $p_{ij}$. Assume we are in i state at time $t$, the transition probability only depends on the state at time $t$. The previous states before time $t$ has no impact on the future. The assumption can be presented as:

$$P(s_t = i | s_1, \ldots, s_{t-1}) = P(s_t = i | s_{t-1}) \tag{3}$$

The probability of an output observation $x_t$ depends only on the i state.

$$P(x_t | s_1, \ldots, s_T, y_1, \ldots, y_{t-1}) = P(y_t = i | s_t = i)$$

Therefore, if there is a autoregressive lag order $p$, the probability of observing $x_t$ based on past $p$ days is :

$$P(x_t = i | s_1, \ldots, s_{t-p}, y_1, \ldots, y_{t-p}; \theta) \tag{4}$$

where $\theta = \{\mu, \beta, \sigma\}$ is the parameters of autoregressive model.

### 3.2.1. Transition matrix

We consider the question of determine the probability from i state at time $t$ to j state at time $t+1$ which is denotes as $a_{ij}$

$$P(s_t|s_{t-1}=i)=a_{i,j}$$

where

$$\sum_{j=1}^{N} a_{i,j}=1 \text{ for } i,j=1,\ldots,N$$

In our paper, we consider two states so $N=2$ and

$$\begin{aligned}
P(s_{t+1}=0|s_t=0)&=a_{00},\\
P(s_{t+1}=0|s_t=1)&=a_{01},\\
P(s_{t+1}=1|s_t=0)&=a_{10},\\
P(s_{t+1}=1|s_t=1)&=a_{11},
\end{aligned} \tag{5}$$

### 3.2.2. Forward probability

Our first task is to maximize the likelihood function of the observed sequence

$$P(x_1,\ldots,x_T|\lambda)$$

Each hidden state only produce one signal. The forward probability $\alpha_t(i)$ is the likelihood of being in $i$ state with $t$ signals at time $t$ which one observed from time 1 to time $t-1$.

$$P(x_1,\ldots,x_t,s_t|\lambda)$$

Daily trading volume is used to initialized the probability of state. The observing signal is stock return during HMM training, so we can investigate the trend of stock price in the given state.

The initial hidden sequence is obtained from 3. Then, for a particular hidden sequence state $S=\{s_1,\ldots,s_T\}$the likelihood of observation sequence is

$$P(X|S)=\prod_{t=1}^{T} P(x_t|s_t) \tag{6}$$

For example, the initial state from time $t=1,2,3$ is $\{0,1,0\}$. We use stock return as the observation signal. The probability of the sequence of $\{rise, fall, rise\}$ for stock return is

$$P(\{high,low,high\}|0,1,0)=P(high|0)\times P(low|1)\times P(high|0)$$

The joint probability of the signal sequence and the state sequence is:

$$P(X,S)=P(X|S)\times P(S) \tag{7}$$

It leads to the total probability of the observations by summing over all possible hidden state sequences that could generate the observation sequence:

$$P(X)=\sum_{t=1}^{T} P(X,S) \tag{8}$$

If we have $N$ hidden states and a signal sequence of $T$ observations, there are $N^T$ possible hidden sequence. To save the computational cost, we use forward algorithm to compute the likelihood of the signal sequence. We build a dynamic programming matrix $\alpha$ which represents the probability of being in state $i$ having observed signals in past $p$ days.

The empty dynamic programming matrix is initialized as:

$$\alpha_1(j) = P(x_1, s_1 = i|\lambda) = \pi_i b_j(x_1) \tag{9}$$

where $\pi_i$ is initialized from signal sequence of trading volume. For example, if we referred 0 as the state $j$, then if we have five "0" signal in the calibration of 10 days, the initial probability $\pi_j = 1/4$.

For computational simplicity, we randomly generated the probability of the initial transition matrix for state and the observation $x_0$ likelihood given the current state $i$ which is denoted as $b_j(x_0)$ to be freely parameterizd at the initial step.

Each forward algorithm trellis $\alpha_t(i)$ represent the probability of being in state $i$ after seeing the first t observations:

$$\alpha_t(i) = P(x_0, \ldots, x_t, s_t|\lambda) \tag{10}$$

In recursion phase,

$$\begin{aligned}\alpha_t(j) = P(x_1, \ldots, x_t, s_t = j|\lambda) &= \sum_{j=1}^{N} P(x_1, \ldots, x_{t-1}, s_{t-1} = i|\lambda) \cdot P(s_t = j|s_{t-1} = i) \cdot P(x_t|s_t = j) \\ &= \sum_{j=1}^{N} \alpha_{t-1}(j) a_{ij} b_j(x_t)\end{aligned} \tag{11}$$

Finally, we get the total probability of observing the sequence of signal :

$$P(X|\lambda) = \sum_{j=1}^{N} P(X, s_t = j|\lambda) = \sum_{j=1}^{N} \alpha_T(i) \tag{12}$$

*3.2.3. Backward probability*

The backward probability of $\beta$ is the probability of observing the signals from time $t+1$ to the end when we are at time $t$ in i state.

$$\beta_t(i) = P(x_{t+1}, x_{t+2}, \ldots, x_T|s_t = i, \lambda) \tag{13}$$

It is computed in a similar manner to the forward probability but in opposite order. The $\beta_T(i)$ is initialized to 1. Then, Given that we are in $i$ state and time $t$

$$\begin{aligned}\beta_t(j) &= \sum_{j=1}^{N} P(x_{t+1}, \ldots, x_T|s_t = i, \lambda) \cdot P(s_{t+1} = j|s_t = i) \cdot P(x_{t+1}|s_{t+1} = j) \\ &= \sum_{i=1}^{N} \beta_{t+1}(j) a_{ij} b_j(x_{t+1})\end{aligned} \tag{14}$$

We obtain the total probability of the sequence of signal in state $i$ at time $t$.

$$\begin{aligned}P(X|\lambda) &= \sum_{i=1}^{N} P(x_2 \ldots x_T|s_1 = i, \lambda) \cdot P(s_1|\lambda) \cdot P(x_1|s_1 = i) \\ &= \sum_{i=1}^{N} \beta_1(i) \pi_i b_j(x_1)\end{aligned} \tag{15}$$

*3.3. Forward-Backward algorithm*

The standard algorithm for HMM training is the forward-backward, or Baum- Welch algorithm which is a special cased of Expectation-Maximization(EM) Algorithm. The section introduce how to use forward and backward probability to re-estimate the parameter $A$ and $B$ for HMM. The goal of forward-backward algorithm is to re-estimate the $A$ and $B$ from the observing sequence of signal and the possible states sequence in HMM.

Let us begin by seeing how to re-estimate observation probability $b_i(x_t)$ given the current state $i$. The probability of being in state $i$ at time $t$:

$$\gamma_t(i) = P(s_t = i|X', \lambda) = \frac{P(X', S' = i|\lambda)P(X'|S' = i, \lambda)}{P(X'|\lambda)} \tag{16}$$

where $S' = \{s_t = i, s_{t+1} = j\}$ and $X' = \{x_1, \ldots, x_t\}$. We can obtain the $\gamma_t$ by following the law of probability

$$\gamma_t(i) = P(s_t = j|X', \lambda) = \frac{P(s_t = i, X'|\lambda)}{P(X'|\lambda)} \tag{17}$$

The probability of a given stock return $x_t$ in a given state $i$ is:

$$P(x_t|S' = i, X', \lambda) = \frac{\sum_{t=1,x_t=1} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} = \frac{\sum_{t=1}^{T} P(x_t|s_t = i)P(s_t = i|X', \lambda)}{\sum_{t=1}^{T} P(s_t = i|X', \lambda)} \tag{18}$$

One can think of (18) as the probability density function(PDF) of the observation likelihood in state $i$. Because the empirical results illustrates that the distribution of the stock returns departures from normal distribution. Therefore, we decided to let the predicted process of stock return in our model be based on an univariate AL distribution. The detail of AL distribution is introduced in appendix.

$$b_i(x_t) = P(x_t|S' = i, X', \lambda) = \frac{1}{\sigma} \frac{\kappa}{1 + \kappa^2} \left\{ \begin{array}{l} exp(-\frac{\kappa}{\sigma}x), \text{ if } x \geq 0 \\ exp(\frac{1}{\sigma\kappa}x), \text{ if } x < 0 \end{array} \right. \tag{19}$$

where

$$\kappa = \frac{2\sigma}{\mu + \sqrt{4\sigma^2 + \mu^2}} \tag{20}$$

We also need the formula for recomputing the transition probability.

The probability of being in $i$ state at time $t$ and $j$ state at time $t + 1$ when observing sequence at time $t$ is:

$$\delta_t(i, j) = P(s_t = i, s_{t+1} = j|X, \lambda) \tag{21}$$

Once again, we can calculate (21) by the law of probability

$$P(s_t = i, s_{t+1} = j|X', \lambda) = \frac{P(s_t = i, s_{t+1} = j, X'|\lambda)}{P(X'|\lambda)} \tag{22}$$

$P(s_t = i, s_{t+1} = j, X'|\lambda)$ can be obtained by

$$\begin{aligned} P(s_t = i, s_{t+1} = j, X'|\lambda) &= \alpha_t(i)\beta_{t+1}(j)a_{ij}b_j(x_{t+1}) \\ &= \sum_{s_{p+1}=1}^{N} \sum_{s_p=1}^{N} \cdots \sum_{s_1=1}^{N} P(s_{p+1}|s_p)a_{i,j}b_j(x_{t+1}) \end{aligned} \tag{23}$$

Given that in state i at the time $t$, the probability of the observing sequence in past $t$ days is

$$P(X'|\lambda) = \sum_{j=1}^{N} P(x_t, s_t = i|\lambda) = \sum_{j=1}^{N} \alpha_t(i)\beta_t(i) \tag{24}$$

Therefore, the final equation for $\delta_t(i,j)$

$$
\begin{aligned}
\delta_t(i,j) &= \frac{P(s_t=i, s_{t+1}=j, x_t|\lambda)}{P(X'|\lambda)} = \frac{P(x_t, s_t=i|\lambda)P(s_t=i, s_{t+1}=j)P(x_t|S', X', \lambda)}{P(X'|\lambda)} \\
&= \frac{\alpha_t(i)\beta_{t+1}(j)a_{ij}b_j(x_{t+1})}{\sum_{j=1}^{N}\alpha_t(i)\beta_t(i)}
\end{aligned}
\tag{25}
$$

To estimate $a_{ij}$, we need the total expected number of transition from state $i$ and the expected number of transition from state $i$ to state $j$. Therefore, the final formula for the element $a_{ij}$ in transition matrix is

$$
\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} P(s_t=i, s_{t+1}=j|X,\lambda)}{\sum_{t=1}^{T-1} P(s_t=i|X,\lambda)} = \frac{\sum_{t=1}^{T-1}\delta_t(i,j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N}\delta_t(i,*)}
\tag{26}
$$

### *3.3.1. Optimization*

The forward-backward algorithm start with the initial estimate of $\lambda = (A, B)$. We use EM algorithm fit the parameters $A$ and $B$ for HMM model such that it can find the maximum likelihood estimation(MLE) of $\lambda$. EM algorithm gives the efficient method for calculating MLE which is consisted with iterative two steps: expectation step(E-step) and maximization step(M-step). In the E-step, we compute the $\gamma$ and $\delta$ from (17) and (25). In the M-step, we com we use $\gamma$ and $\delta$ to update the $A$ and $B$. Our convergence criterion is to stop when the maximal element of $|\lambda_{\ell+1} - \lambda_\ell|$ is less than $10^{-5}$. If the convergence criterion indicates that the parameters did not converge, we calculated the new state from the sequence of $P(s_t = i|y_{t-1})$.

We followed the proof from Obervation 1 in [**?** ] that the EM algorithm serve as analytic solution to the optimizing problem for HMM.

**Theorem 3.1.** *Let $\hat{\lambda}_\ell$ denote the estimate of parameter of vector from previous iteration where $\ell$ is the number of iteration. $\hat{\lambda}_{\ell+1}$ is associated with higher value of the likelihood function $\hat{\lambda}_\ell$ such that*

$$
P(X|\hat{\lambda}_{\ell+1}) > P(X|\hat{\lambda}_\ell)
$$

*with the equality only if $\hat{\lambda}_{\ell+1} = \hat{\lambda}_\ell$*

*Proof.* Let $Q(\lambda_{\ell+1}|\lambda_\ell, X)$ denote the expected log-likelihood estimate for $\lambda_{\ell+1}$. The expectation is based on the distribution of $\lambda_\ell$:

$$
Q(\lambda_{\ell+1}|\lambda_\ell, X) = P(X, S|\lambda_\ell)\int_S \log P(S, X|\lambda_{\ell+1})
\tag{27}
$$

$$
\begin{aligned}
& Q(\lambda_{\ell+1}|\lambda_\ell, X) - Q(\lambda_\ell|\lambda_\ell, X) \\
&= P(X, S|\lambda_\ell)\int_S \log\left[\frac{P(S, X|\hat{\lambda}_{\ell+1})}{P(S, X|\hat{\lambda}_\ell)}\right] \\
&\leq P(X, S|\lambda_\ell)\int_S \log\left[\frac{P(S, X|\hat{\lambda}_{\ell+1})}{P(S, X|\hat{\lambda}_\ell)}\right] \\
&= \int_S \log\left[P(S, X|\hat{\lambda}_{\ell+1}) - P(S, X|\hat{\lambda}_\ell) - 1\right] \\
&= P(X|\hat{\lambda}_{\ell+1}) - P(X|\hat{\lambda}_\ell)
\end{aligned}
\tag{28}
$$

Thus, if

$$
Q(\lambda_{\ell+1}|\lambda_\ell, X) > Q(\lambda_\ell|\lambda_\ell, X)
$$

then

$$
P(X|\hat{\lambda}_{\ell+1}) > P(X|\hat{\lambda}_\ell)
$$

□

Therefore, it proves that the $\{\hat{\lambda}_{\ell}{}_{\ell=1}^{\infty}\}$ converges to the local maximum value estimation(MLE).

### 3.4. Parameters estimation

After determining stock return to the state, we maximize the likelihood function with respect to the AR(1) parameters. Therefore, we consider the case for stock return $y_1, \ldots, y_n$ is i.i.d. Gaussian distribution with mean vector and covariance matrix in the state $i$:

$$P(X_t|S_t = i)$$

All the parameters in autoregressive model is presumed to shift with the state $s_t$. Therefore, we consider conditional maximum likelihood estimation of a scalar m-th order autoregression.

In asymmetric Laplace distribution, Kozubowski et al. [24] showed that the conditional probability distribution of $y_{t+1}$ given $Y = y_t$ is:

$$P(y_2|Y = y_1) = \frac{\xi^\lambda \exp(\beta'(y_{t+1} - \mu))K_{k/2-\lambda}\left(\alpha\sqrt{\delta^2 + (y_{t+1} - \mu)'\Delta^{-1}(y_{t+1} - \mu)}\right)}{(2\pi)^{k/2}|\Delta|^{1/2}\delta^\lambda K_\lambda(\delta\xi)[\sqrt{\delta^2 + (y_{t+1} - \mu)'\Delta^{-1}(y_{t+1} - \mu)}/\alpha]^{k/2-\lambda}} \tag{29}$$

where $K$ is the Bessel function and $\chi = \sqrt{\alpha^2 + \beta^2}$. For $v = n + \frac{1}{2}$ and $n = 0, 1, 2, \ldots$ the Bassel function $K_v$ can be expressed explicitly as

$$K_{n+1/2}(x) = \sqrt{\pi/2}x^{-1/2}e^{-x}\left(1 + \sum_{i=1}^{n} \frac{(n+i)!}{(n-i)!i!}(2x)^{-i}\right)$$

For simplicity, set $\delta = 1$ and $\lambda = 1$, (29) become

$$\frac{\exp(\beta'(\phi(y_{t+1} - \mu)))K_{\frac{k}{2}}(\alpha\sqrt{2\Delta^{-1}\phi(y_{t+1} - \mu)})}{(2\pi)^{k/2}|\Delta|^{1/2}K_1(0)[\sqrt{\Delta^{-1}(\phi(y_t - \mu))^2}/\alpha]^{k/2-1}} \propto \frac{\chi}{2\alpha K_1(x)}\exp(-\alpha\sqrt{1 + \phi^2(y_t - \mu)^2} + \beta\phi(y_t - \mu)) \tag{30}$$

where

$$y_{t+1} - \mu_{s_t} = \phi(y_t - \mu_{s_{t-1}}) + \varepsilon_t$$

The log likelihood of $P(y_2|Y = y_1)$ is

$$\log P(y_2|Y = y_1) = -\alpha\sqrt{1 + \phi^2(y_t - \mu)^2} + \beta\phi(y_t - \mu) \tag{31}$$

We aim to find the autoregressive parameter which can maximize the likelihood of $P(y_2|Y = y_1)$. Therefore,

$$\frac{\partial \log P}{\partial \phi} = -\frac{\alpha}{2}\frac{2\phi(y_t - \mu)^2}{\sqrt{1 + \phi^2(y - \mu)}} + \beta(y - \mu) = 0 \tag{32}$$

$$\frac{\phi(y_t - \mu)^2}{\sqrt{1 + \phi^2(x - \mu)^2}} = \frac{\beta(y - \mu)}{\alpha} \Rightarrow \frac{\phi^2(y_t - \mu)^4}{1 + \phi^2(x - \mu)^2} = \frac{\beta^2(y - \mu)^2}{\alpha^2} \tag{33}$$

$$\alpha^2\phi^2(y - \mu)^4 = \beta^2(y - \mu)^2 + \beta^2(y - \mu)^2\phi^2(y - \mu)^2 \tag{34}$$

$$\begin{aligned} &\phi^2[(y - \mu)^4\alpha^2 - \beta^2(y - \mu)^4] = \beta^2(y - \mu)^2(y - \mu) \\ &\phi^2[(y - \mu)^4(\alpha^2 - \beta^2)] = \beta^2(y - \mu)^2 \end{aligned} \tag{35}$$

If $\alpha^2 - \beta^2 = 0$, the formulas of generalized hyperbolic distribution can be interpreted as the corresponding limit expressions by means of the asymptotic relation.

Therefore, if we set $\alpha^2 - \beta^2 = 0$, we get

$$\beta^2(y_t - \mu)^2 = 0 \tag{36}$$

If we consider the case for $y - \mu \neq 0$, then $\beta = 0$. When $\lambda = 1$, $\delta = 0$ and $\beta = 0$, the distribution become symmetric scaled Laplace distribution. Therefore, the maximum likelihood of conditional distribution appear when $p(y_2|Y = y_1)$ is symmetric Laplace distribution.

Then, $Y$ could be generated as a multivariate scale mixture of Gaussians according to

$$Y = \mu + \sqrt{Z}\Sigma^{1/2}X \tag{37}$$

where $Z$ is the standard exponential variate and $X$ follows the multivariate Gaussian distribution $N(0, \Sigma)$ independent of $Z$.

Given the state j, the multivariate Gaussian with pdf given as

$$p_{Y|Z} = \frac{1}{\sqrt{2\pi z}} \exp\left[-\frac{1}{2z}(y - \mu_j)^t\Sigma^{-1}(y - \mu_j)\right] \tag{38}$$

$$\log p_{Y|Z}(\mu, \Sigma) \propto -\frac{1}{2}\sum_{t=1}^{N}\frac{1}{z_t}(y_t - \mu)\Sigma^{-1}(y_t - \mu)' \tag{39}$$

The mean is obtained from (37),

$$E(Y) = \mu Y = \mu \tag{40}$$

Maximizing the log-likelihood function with regard to $\mu_j$ and $\Sigma_j$, we obtain

$$\frac{\partial \log p_{Y|Z}(\mu, \Sigma)}{\partial \mu_j} = -\frac{1}{2}\cdot 2\sum_{t=1}^{T} -\frac{1}{z_t}\Sigma^{-1}(y_t - \mu_j) = \sum_{t=1}^{T}\frac{1}{z_t}\Sigma^{-1}(y_t - \mu_j) \tag{41}$$

and

$$\frac{\partial \log p_{Y|Z}(\mu, \Sigma)}{\partial \Sigma^{-1}} = \sum_{t=1}^{T}\frac{1}{2}\Sigma_j - \frac{1}{2}(y_t - \mu_j)(y_t - \mu_j)' \tag{42}$$

where $\Sigma$ is the matrix whose$(i, j)$ entry is the covariance.

$$\Sigma = \begin{bmatrix} E[(y_1 - E[y_1])(y_1 - E[y_1])] & \dots & E[(y_1 - E[y_1])(y_n - E[y_n])] \\ E[(y_2 - E[y_2])(y_1 - E[y_1])] & \dots & E[(y_2 - E[y_2])(y_n - E[y_n])] \\ \vdots & \ddots & \vdots \\ E[(y_n - E[y_n])(y_1 - E[y_1])] & \dots & E[(y_n - E[y_n])(y_n - E[y_n])] \end{bmatrix} \tag{43}$$

Solving (42) and (41) for $\mu^{(\ell+1)}$ and $\Sigma^{(\ell+1)}$, the EM equations are thus

$$\mu_j^{(\ell+1)} = \frac{\sum_{t=1}^{T} y_t \frac{1}{z_t} P(s_t = j|X, \lambda)}{\sum_{t=1}^{T}\frac{1}{z_t}P(s_t = j|X, \lambda)} \tag{44}$$

$$\Sigma_j^{(\ell+1)} = \frac{\sum_{t=1}^{T}\frac{1}{z_t}(y_t - \mu_j)(y_t - \mu_j)'P(s_t = j|X, \lambda)}{\sum_{t=1}^{T}\frac{1}{z_t}P(s_t = j|X, \lambda)} \tag{45}$$

Because the probability of the next state depends on the current state and its transition probability, the Markov process is very suitable to incorporated with $AR(1)$ model. In the case of ARHMM, the principles of calculating the transition probability are essentially the same in calibration in HMM.

*3.4.1. Prediction of the process*

After getting the covariance matrix from (42), we now want to measure the linear correlation between the current day and the next day such that we could obtain the coefficient $\phi$ for ARHMM.

Estimating stock price is achieved using the ARHMM(1) model that can be expressed as a linear combination of lagged values of the noise series,

$$y_t = \mu_i + \beta * y_{t-1} + \varepsilon \tag{46}$$

where

$$\varepsilon = \frac{\sum_{t=1}^{T} e_t^x}{T}$$

and $e$ denotes the square of the difference between the real price and the theoretical price and $w$ is the weight of the state from **??**.

$$e = y_t - \mu_i + \beta * y_{t-1}$$

To obtain the autoregressive parameters $\beta^{(i)}$ in state $i$, we calculate the correlation coefficient between the mean value of stock return in state $i$ and the stock return at time $t$

$$\begin{aligned} corr(Y) &= \begin{bmatrix} 1 & \frac{E[(y_1-E[y_1])(y_1-E[y_1])]}{\sigma(y_1)\sigma(y_2)} & \cdots & \frac{E[(y_1-E[y_1])(y_n-E[y_n])]}{\sigma(y_1)\sigma(y_n)} \\ \frac{E[(y_2-E[y_2])(y_1-E[y_1])]}{\sigma(y_2)\sigma(y_1)} & 1 & \cdots & \frac{E[(y_2-E[y_2])(y_n-E[y_n])]}{\sigma(y_2)\sigma(y_n)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{E[(y_n-E[y_n])(y_1-E[y_1])]}{\sigma(y_n)\sigma(y_1)} & \cdots & \cdots & 1 \end{bmatrix} \\ &= (diag(\Sigma))^{-\frac{1}{2}}\Sigma(diag(\Sigma))^{-\frac{1}{2}} \end{aligned} \tag{47}$$

For the covariance of each state, we collected all the stock prices in the calibration and calculated the theoretical price. Given a pair $(y_{t+1}, y_t)$, each element in correlation matrix is

$$corr(y_{t+1}, y_t) = \rho_{y_{t+1},y_t} = \frac{Cov(y_{t+1}, y_t)}{Var(y_{t+1})Var(y_t)} \tag{48}$$

$$\begin{aligned} Cov(y_{t+1}, y_t) &= E(y_{t+1}, y_t) - E(y_{t+1})E(y_t) \\ &= E((\beta y_t + \varepsilon)y_t) - E(y_{t+1})E(y_t) \\ &= \beta E(y_t^2) + E(\varepsilon y_t) - E(y_{t+1})E(y_t) \\ &= \beta Var(y_t) + E(\varepsilon y_t) - E(y_{t+1})E(y_t) \\ &= \beta Var(y_t) - E(y_{t+1})E(y_t) \end{aligned} \tag{49}$$

$$\begin{aligned} Cov(y_t, y_{t-1}) &= E(y_t, y_{t-1}) - E(y_t)E(y_{t-1}) \\ &= \beta Var(y_{t-1}) - E(y_t)E(y_{t-1}) \end{aligned} \tag{50}$$

Since we assume 2nd-order stationary,

$$Var(y_t) = Var(y_{t-1}) \tag{51}$$

so

$$Var(y_{t+1}) = Var(y_t) \tag{52}$$

$$corr(y_{t+1}, y_t) = \rho_{y_{t+1},y_t} = \frac{Cov(y_{t+1}, y_t)}{Var(y_{t+1})Var(y_t)} = \frac{E[(Y_t - \mu_j)(Y_{t-1}\mu_j)]}{Var(y_{t+1})Var(y_t)} \tag{53}$$

$$
\begin{aligned}
\phi^{(i)} &= \frac{Cov(y_t, y_{t-1})}{\sigma^2} \\
&= \frac{\sum_{X_l=i}(y_{t,j} - \mu_j^{(X_t)})(y_{t-1,j} - \mu_j^{(X_{t-1})})}{\sum_{X_l=i}(y_{t,j} - \mu_j^{(X_t)})^2}
\end{aligned}
\tag{54}
$$

where $\mu$ is the mean value of the stock return for the calibration in $i$ state.

Suppose we have a sample of size $p$ from a vector-valued process($y_t \in \mathbb{R}^p$). For example, we can predict the value for the next time step $(t+1)$ given the observations at the last two time steps ($t-1$ and $t-2$). As a regression model, this would look like the following:

$$
y = \phi X_t + \varepsilon
$$

where $e$ is the white noise.

Then, we multiply the current price with likelihood stock return and add the value to the current price. Therefore, the next day's stock price forecast is established by adding above value to the current day's closing price.

## 4. Application to specific stock forecasting

### *4.1. Method*

In this section we present empirical study to assess the quality of ARHMM with AL innovations. In our empirical work, we first derived the stationary distribution of the process and cluster trading volume with K-means clustering, to initialize the state and transition for each group. To transform the time series data into stationary process, we used stock returns to model the financial data to maintain a consistent mean and variance.

If the difference of trading volume between previous day is positive, we initialized the date into "0" state; if it is negative, we initialized the date into "1" state. We considered $y_t$ a vector process in which $s_t$ follows a two-state, first order Markov process with a time lag of 20 days. The calculation of means and mean squared errors is based on 20 days. Combined the trained HMM and AR model, we adjust the AR(1) model and obtain the likelihood probability of state sequence using Eq. **??**. The parameters in ARHMM(1) is shifted by the maximum likelihood estimation of state $i$ in time $t$ under AL distribution. Because the state path updates during each iteration, we adjusted our pdf and re-estimate the parameters $\lambda = (\pi, A, f_X)$ using forward-backward algorithm and AR parameter $\beta$. After the transition probability is convergent, the state path stops to updating. Then, we obtained the $\beta$ for the current state which lets us estimate the stock price on the next day. For example, if ARHMM predicts the state of the next day is increasing which is indicated as 1, we estimated the stock price as

$$
\triangle y_{t+1} = \mu_x + \beta_2(y_t - \mu_x) + \varepsilon
$$

$$
y_{t+1} = y_t + \beta_2 \triangle y_{t+1}
$$

where $\triangle y_{t+1}$ is the predicted stock return for the next day.

We implemented another experiments that using the change of stock return to be the signal of ARHMM.

Figure 4 shows the predicted and the actual prices of stock from 5/24/2018 to 5/24/2019. Our convergence criterion is to stop when the maximum element of $|\lambda_{l+1} - \lambda_l|$ is less than $10^{-8}$ where $\lambda$ are the unknown parameters to be estimated in a single vector.

The experiment was evaluated by calculating Mean Absolute Percent Error(MAPE) and the normalized root mean square error (NRMSE). The MAPE is the average of the ratio of the actural price to the forecast price.

$$\text{MAPE} = (\frac{1}{n}\sum \frac{| y_\text{A} - y_\text{F} |}{y_\text{A}}) * 100\%$$

where $y_\text{A}$ is the actual stock price and $y_\text{F}$ is the forecast price.
The root mean square error (RMSE) is the square root of the difference between forecast value and real value.

$$\text{RMSE} = \sqrt{E((y_\text{A} - y_\text{F})^2)} = \sqrt{\frac{\sum_{t=1}^{T}(y_\text{A} - y_\text{F})^2}{T}}$$

Since the values of error between actual price and predicted price is affected by the scale of stock price, we normalized the RMSE by the range of the price to facilitate the comparisons between datasets with different scale.

$$\text{NRMSE} = \frac{\text{RMSE}}{y_\text{max} - y_\text{min}}$$

For example, because AMZN and GOOG price scales is larger than other three stocks in our experiment, their RMSE value is larger than 10.

We presented a case study to illustrate the proposed methodology. Our results showed that observing the trading volume to predict the trend in the stock price on the next day is more accurate than observing the stock change on the current day. In other words, observing the daily tranding volume to predict the trend of stock price in next day is more accurate than observe the stock change in the current day.

Table 1 summarized the evidence if model observes stock return as signal. Table 2 summarized the evidence if our model observes volume as signal. Before we involve the volume to the model, MAPE and NRMSE is lower than 2%. After we use the trading volume to be the signal for model, the verification measure all decrease. Some stock's MAPE and NRMSE even lower than 1%. The mean and maximum error also decrease. It shows that the prediction result via our approach is more accurate than past researches [ [20], [35]].

| Stock Name | MAPE | NRMSE | Mean Error |
|---|---|---|---|
| FB | 0.015 | 0.035 | 2.346 |
| AAPL | 0.012 | 0.024 | 1.858 |
| GOOGL | 0.013 | 0.033 | 12.518 |
| AMZN | 0.012 | 0.021 | 16.560 |
| AMJ | 0.013 | 0.045 | 0.375 |

Table 1: Prediction accuracy of stock when the stock return is the signal for ARHMM model in the period 5/24/2018-5/29/2019.

## 5. Conclusion

In this paper, we proposed a ARHMM based on the shift of heterogeneous beliefs in the market. In the equilibrium, individual investors exhibit the disposition effect. However, if the overconfidence take over the market, it increase the trading volume and market depth.

| Stock Name | MAPE | NRMSE | Mean Error |
|---|---|---|---|
| FB | 0.010 | 0.026 | 1.645 |
| AAPL | 0.009 | 0.020 | 1.416 |
| GOOGL | 0.009 | 0.025 | 9.12 |
| AMZN | 0.011 | 0.016 | 12.357 |
| AMJ | 0.009 | 0.034 | 0.275 |

Table 2: Prediction accuracy of stock when daily trading volume is the signal for ARHMM model in the peroid 5/24/2018-5/29/2019.

We developed the ARHMM based tool such that the stock return forecasting is based on the degree of heterogeneous belief in the market. For simplicity, we consider two features for a stock: the quantity of increasing and decreasing for trading volume. The next day's closing price is taken as the target price associate with the two features. Our objective is to predict the next day's stock price for a specific stock using aforementioned ARHMM model. The model has capability to infer the volatility in the market via measuring the change of daily trading volume. We use K-Mean Algorithm to implement the signal recognition and initialize the state for a sequence of time series data. Then, we present the estimation and optimizing procedure for the ARHMM. The parameter is estimated under AL distribution, so the model can be useful for modeling time series data such as stock return that tend to be peaked, skewed. To illustrate the proposed strategy, we select five stocks daily return series to verify our model. The empirical results suggested that using daily trading volume as the signal for ARHMM is a good tool for forecasting the next day price. Through our experiments, it also illustrates that ARHMM with AL noise is useful for modeling asymmetry and steep peak data.

i

## References


[1] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, G. Amaratunga, Ensemble deep learning for regression and time series forecasting, in: 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL), 2014, pp. 1–6. doi:10.1109/CIEL.2014.7015739.

[2] Z. Kilimci, A. Okay Akyuz, M. Uysal, S. Akyokus, M. Ozan Uysal, B. Atak Bulbul, M. Ali Ekmis, An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain, Complexity 2019 (2019) 1–15. doi:10.1155/2019/9067367.

[3] M. Gupta, A. Asthana, N. Joshi, P. Mehndiratta, Improving time series forecasting using mathematical and deep learning models, in: A. Mondal, H. Gupta, J. Srivastava, P. K. Reddy, D. Somayajulu (Eds.), Big Data Analytics, Springer International Publishing, Cham, 2018, pp. 115–125.

[4] D. Sahoo, Q. Pham, J. Lu, S. C. H. Hoi, Online deep learning: Learning deep neural networks on the fly, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 2660–2666. doi:10.24963/ijcai.2018/369. URL https://doi.org/10.24963/ijcai.2018/369

[5] B. R. Massimo Caccia, Option pricing and hedging for discrete time autoregressive hidden markov model, in: eprint arXiv:1707.02019, 2017, dOI = 10.2139/ssrn.2995944.

[6] R. J. Povinelli, Identifying temporal patterns for characterization and prediction of financial time series events, in: Proceedings of the First International Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining-Revised Papers, TSDM '00, Springer-Verlag, London, UK, UK, 2001, pp. 46–61.
URL http://dl.acm.org/citation.cfm?id=645905.673094

[7] Z. GHAHRAMANI, An introduction to hidden markov models and bayesian networks, International Journal of Pattern Recognition and Artificial Intelligence 15 (01) (2001) 9–42. arXiv:https://doi.org/10.1142/S0218001401000836, doi:10.1142/S0218001401000836.
URL https://doi.org/10.1142/S0218001401000836

[8] C. Li, G. Biswas, Temporal pattern generation using hidden markov model based unsupervised classification, in: D. J. Hand, J. N. Kok, M. R. Berthold (Eds.), Advances in Intelligent Data Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999, pp. 245–256.
URL https://doi.org/10.1007/3-540-48412-4_21

[9] N. Nguyen, An analysis and implementation of the hidden markov model to technology stock prediction, in: Risks, Vol. 5, 2017, p. 62, dOI= https://doi.org/10.3390/risks5040062.

[10] S. H. Lihn, Hidden markov model for financial time series and its application to s&p 500 index, in: Quantitative Finance, 2017, available at SSRN: https://ssrn.com/abstract=2979516.

[11] T. Odean, Volume, volatility, price, and profit when all traders are above average, The Journal of Finance 53 (6) (1998) 1887–1934. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00078, doi:10.1111/0022-1082.00078.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/0022-1082.00078

[12] A. ATMAZ, S. BASAK, Belief dispersion in the stock market, The Journal of Finance 73 (3) 1225–1279. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12618, doi:10.1111/jofi.12618.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12618

[13] C. Shenoy, P. P. Shenoy, Modeling Financial Portfolios Using Belief Functions, Physica-Verlag HD, Heidelberg, 2002, pp. 316–332. doi:10.1007/978-3-7908-1798-0_11.
URL https://doi.org/10.1007/978-3-7908-1798-0_11

[14] M. F. Hellwig, On the aggregation of information in competitive markets, Journal of Economic Theory 22 (3) (1980) 477 – 498. doi:https://doi.org/10.1016/0022-0531(80)90056-3.
URL http://www.sciencedirect.com/science/article/pii/0022053180900563

[15] D. W. Diamond, R. E. Verrecchia, Information aggregation in a noisy rational expectations economy, Journal of Financial Economics 9 (3) (1981) 221 – 235. doi:https://doi.org/10.1016/0304-405X(81)90026-X.
URL http://www.sciencedirect.com/science/article/pii/0304405X8190026X

[16] J. Y. Campbell, S. J. Grossman, J. Wang, Trading volume and serial correlation in stock returns, Working Paper 4193, National Bureau of Economic Research (October 1992). doi:10.3386/w4193.
URL http://www.nber.org/papers/w4193

[17] W. S. Area, Relationship between trading volume and security prices and returns, 2005.

[18] J. Scheinkman, W. Xiong, Overconfidence and speculative bubbles, Journal of Political Economy 111 (6) (2003) 1183–1220. arXiv:https://doi.org/10.1086/378531, doi:10.1086/378531.
URL https://doi.org/10.1086/378531

[19] J. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, in: Econometrica, Vol. 57, 1989, pp. 357–384, dOI= 10.2307/1912559.

[20] Shanming Shi, A. S. Weigend, Taking time seriously: hidden markov experts applied to financial engineering, in: Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr), 1997, pp. 244–252. doi:10.1109/CIFER.1997.618944.

[21] F. C. Mills, The Behavior of Prices, National Bureau of Economic Research, Inc, 1927.
URL https://EconPapers.repec.org/RePEc:nbr:nberbk:mill27-1

[22] B. Mandelbrot, The Variation of Certain Speculative Prices, The Journal of Business 36 (1963) 394–394. doi:10.1086/294632.
URL https://ideas.repec.org/a/ucp/jnlbus/v36y1963p394.html

[23] S. M. Markus Haas, M. S. Paolella, Modeling and predicting market risk with laplace-gaussian mixture distributions, in: Applied Financial Economics, Vol. 16, 2006, pp. 1145–1162, dOI = 10.1080/09603100500438817.

[24] S. Kotz, T. J. Kozubowski, K. Podgórski, Asymmetric Laplace Distributions, Birkhäuser Boston, Boston, MA, 2001, pp. 133–178. doi:10.1007/978-1-4612-0173-1_3.
URL https://doi.org/10.1007/978-1-4612-0173-1_3

[25] T. Kozubowski, K. Podgorski, Asymmetric laplace laws and modeling financial data, Mathematical and Computer Modelling 34 (2001) 1003–1021. doi:10.1016/S0895-7177(01)00114-5.

[26] A. Karunakaran Jayakumar, A time-series model using asymmetric laplace distribution, in: Statistics & Probability Letters, Vol. 77, 2007, pp. 1636–1640, dOI=https://doi.org/10.1016/j.spl.2005.10.028.

[27] A. A. Trindade, Y. Zhu, B. Andrews, Time series models with asymmetric laplace innovations, Journal of Statistical Computation and Simulation 80 (12) (2010) 1317–1333, dOI=10.1080/00949650903071088.

[28] Z. Trindade, A. A., Approximating the distributions of estimators of financial risk under an asymmetric laplace law, in: Computational Statistics & Data Analysis, Vol. 51, 2007, pp. 3433–3447, available at=10.1016/j.csda.2006.08.004.

[29] M. Linden, A model for stock return distribution, International Journal of Finance & Economics 6 (2) (2001) 159–169. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijfe.149, doi:10.1002/ijfe.149.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ijfe.149

[30] C. M. Lee, B. Swaminathan, Price momentum and trading volume, The Journal of Finance 55 (5) (2000) 2017–2069. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00280, doi:10.1111/0022-1082.00280.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/0022-1082.00280

[31] G.-m. Chen, M. Firth, O. M. Rui, The dynamic relation between stock returns, trading volume, and volatility, Financial Review 36 (3) (2001) 153–174. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6288.2001.tb00024.x, doi:10.1111/j.1540-6288.2001.tb00024.x.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6288.2001.tb00024.x

[32] M. Bouattour, R. Benkraiem, A. Miloudi, Relationships between trading volume, stock returns and volatility: Evidence from the french stock market, 2015.

[33] P. C, The relationship between trading volume and stock returns, Journal of Competitiveness 2011.

[34] D. Jurafsky, J. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Vol. 2, 2008.

[35] M. Hassan, B. Nath, Stockmarket forecasting using hidden markov model: A new approach., in: Proceedings - 5th International Conference on Intelligent Systems Design and Applications 2005, ISDA '05, 2005, pp. 192 – 196.

## Appendix A. The model of price-taking traders

The economy has two assets, a riskless asset and risky asset. There are four time periods. There are N traders($n = 1, \ldots, N$) in the market. A riskless asset and risky asset are exchange in the first rounds of trading time. In the forth period, there is no trading, consume take place at which time the riskless asset pays 1 unit per share and each share of the risky asset pays $\bar{v}$.

At the trading round $1, 2, 3$, the return of risky asset is not known each trader has the prior belief about $\bar{v}$ under Normal distribution with mean $\mu_0$ and precision $h_0$. The signal observed by the investors is $\bar{s}$. The trader $n$ demands for the riskless asset and risky asset are $f_{tn}$ and $x_{tn}$. $\bar{x}$ is the fixed per capita supply of the risky asset. $P_t$ is the price of the risky asset in the trading rounds $1, 2, 3$. Trader n's wealth is

$$\begin{aligned} W_{tn} &= f_{tn} + P_t x_{tn}, \text{ for } t = 1, 2, 3 \\ W_{4n} &= f_{3n} + \bar{v} x_{3n}, \text{ for } t = 4 \end{aligned} \tag{A.1}$$

There is no signal prior to the first round of trading at $t = 1$. Prior to trading at $t = 1, 2, 3$, the trader $n$ receives one of $M$ private signals

$$\bar{s}_{tn} = \bar{v} + \bar{\varepsilon}_{tm} \tag{A.2}$$

where $\varepsilon \sim N(0, h_0)$ Each signal is received by the same number of traders. The assumptions that there are $M < N$ signals in any time is motivated that the number of trader is larger, there are likely to be fewer pieces of information about an asset than there are traders. The trader's posterior is more precise than the rational trader if, after receiving both of her signals

$$\eta h_v + 2(\kappa + (M - 1)\gamma) h_\varepsilon \leq h_v + 2M h_\varepsilon$$

Each trader know that there are $N/M - 1$ other traders receive the same two signals as she is. She believes the precision of these two signals is

$$\kappa h_\varepsilon, \kappa \geq 1$$

and the precision of other $2M-2$ signal is

$$\gamma h_{\varepsilon}, \gamma \leq 1$$

All trader belief the precision of $\bar{v}$ is

$$\eta h_v, \eta \leq 1$$

Let $\Phi_{tn}$ denote the information available to trader $n$ at time $t$. At trading round of time $1, 2, 3$,

$$\begin{aligned}\Phi_{1n} &= \{\} \\ \Phi_{2n} &= [s_{2n} \quad P_2]^T \\ \Phi_{3n} &= [s_{2n} \quad s_{3n} \quad P_2 \quad P_3]^T\end{aligned} \tag{A.3}$$

Each trader has a negative exponential utility function for wealth $w$ with the risk-aversion coefficient of $r$.

$$U(w) = -exp(-aW_v) \tag{A.4}$$

Trades are assume to be myopic, so they only look one period ahead when solving their trading problem. Thus, at time $t = 1, 2, 3$, the maximum equilibrium conditional utility is

$$\max_{x_{tn}} E[-exp(-aW_v)|\Phi_{tn}] \text{ subject to } P_t x_{tn} + f_{tn} \leq P_t x_{t-1i} + f_{t-1i} \tag{A.5}$$

Equilibrium is attained because trader believe they believe that they are behave optimally even though they are not. When solving the maximized problem, trader conjecture that prices are linear function of the average signals

$$\begin{aligned}P_3 &= \alpha_{31} + \alpha_{32}\bar{Y}_2 + \alpha_{33}\bar{Y}_3 \\ P_2 &= \alpha_{21} + \alpha_{22}\bar{Y}_2\end{aligned} \tag{A.6}$$

where

$$\bar{Y}_t = \sum_{n=1}^{N} \frac{s_{tn}}{N} = \sum_{n=1}^{M} \frac{s_{tn}}{M} \tag{A.7}$$

## Appendix B. AL distribution

These facts warrant a new approach to the problem of stock price estimation. We gave an explicit expression [24] for each group's probability density function(pdf) with an AL distribution and discussed the prediction problem.

**Definition Appendix B.1.** *Let $v_k$ be a geometrically distributed random variables with mean $1/p$ where $k \in (0,1)$ The distribution of random variable $Y$ has AL distribution if and only if the geometric compound*

$$X_1 + \cdots + X_{v_p} \tag{B.1}$$

*where identically distribution(i.i.d.) $X$ have finite variance and converge to $Y$ as $k \to 0$. Let $\mu \in \mathbb{R}$ and $\sigma \leq 0$ $Y$ has the following characteristic function:*

$$\psi(t) = \left(1 + \sigma^2 t^2 - i\mu t\right)^{-1} \tag{B.2}$$

*We denote such a random variable as $y_{\sigma,\mu}$ and its distribution as $AL(\sigma, \mu)$ and write $y_{\sigma,\mu} \sim AL(\sigma, \mu)$*

The stock return variable $\{y\}$ is a fat tailed distribution [27] with a location parameter $\theta$, a scale parameter $\tau > 0$ and a skewness parameter $\kappa > 0$, $Y \sim AL(\theta, \kappa, \tau)$. Its PDF is expressed as [24]:

$$f_{X_t^i}(y_t|y_{t-1}) = \frac{1}{\sigma}\frac{\kappa}{1+\kappa^2} \begin{cases} exp(-\frac{\kappa}{\sigma}x), if x \geq 0 \\ exp(\frac{1}{\sigma\kappa}x), if x < 0 \end{cases} \tag{B.3}$$

where

$$\kappa = \frac{2\sigma}{\mu + \sqrt{4\sigma^2 + \mu^2}} \tag{B.4}$$

Values of $\kappa$ in the interval$(0, 1)$ and $(0, \infty)$, correspond to positive (right) skewness and negative (left) skewness, respectively. $X_t$ denotes the adjusted kurtosis of an AL distribution that varies between 3 (the smallest value for the symmetric Laplace distribution when $\kappa = 1$) and 6 (the largest value attained for the limiting exponential distribution as $\kappa \to 0$).

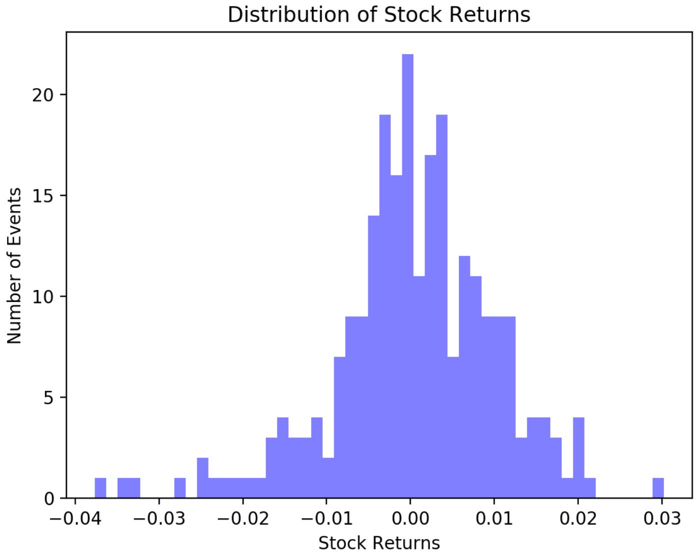
We used the approach from Trindade et al.[27] to adjust the ARHMM with AL noise and furthermore combined it with HMM and changes in the state.

**Definition Appendix B.2.** *(ARHMM model with AL noise) Let* $z_t \overset{iid}{\sim} \mathrm{AL}(\theta, \kappa, \tau)$ *with* $\theta = -\tau(\kappa^{-1} - \kappa)/\sqrt{2}$*. Then,* $\{X_i\}$ *is an ARHMM driven by AL noise if it is a stationary solution of the equations:*

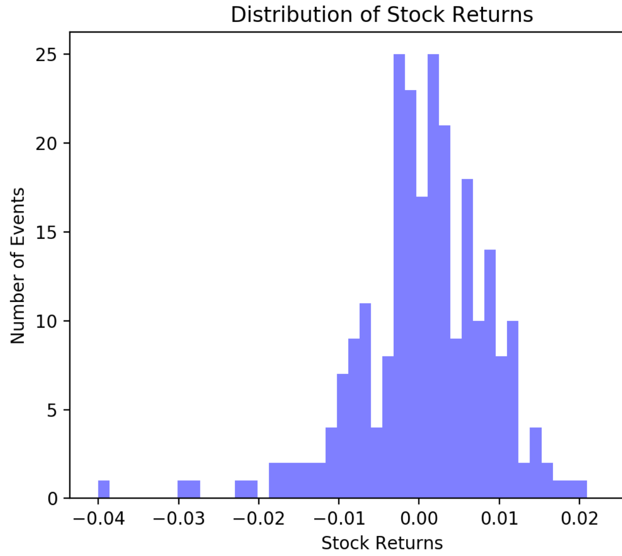
$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \lambda_1 z_{t-1} + \cdots + \lambda_q z_{t-q} + z_t$$

*where the polynomials* $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ *and* $\lambda(z) = 1 + \rho_1 z - \cdots - \rho_q z^q$ *have no common factors. The value of* $\phi$ *and* $\rho$ *is based on current state.*

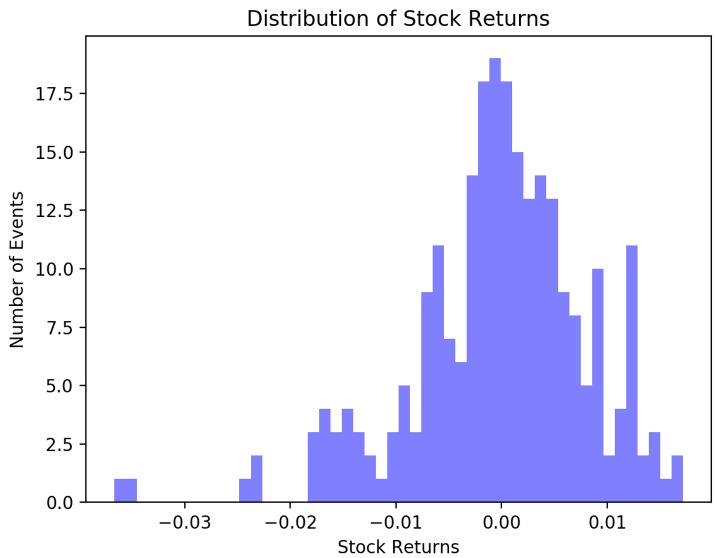
The notation of $f(z_t; \kappa, \tau)$ denotes the pdf of $z_t$ as in Definition 3.1. The model predicts the state by multiplying the probability of current state and transition probability. Then, we used the probability distribution of the stock return to adjust the parameters. We integrate the pdf under the state distribution [17] to obtain the comparative probability in the interval. If the current value is closer to the mean value in the $i$th state, it will have higher probability density in the $i$th state. To optimize the parameters, we iterated the process until the number of iteration is more than 10000 and the transition probability $A$ has converged.



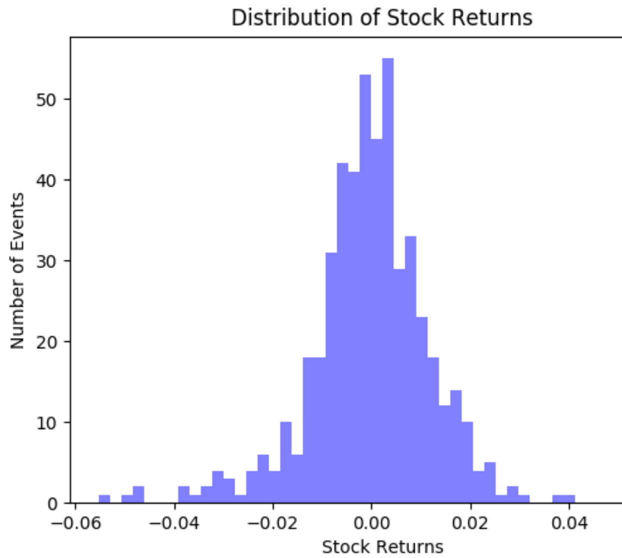

(a) FB



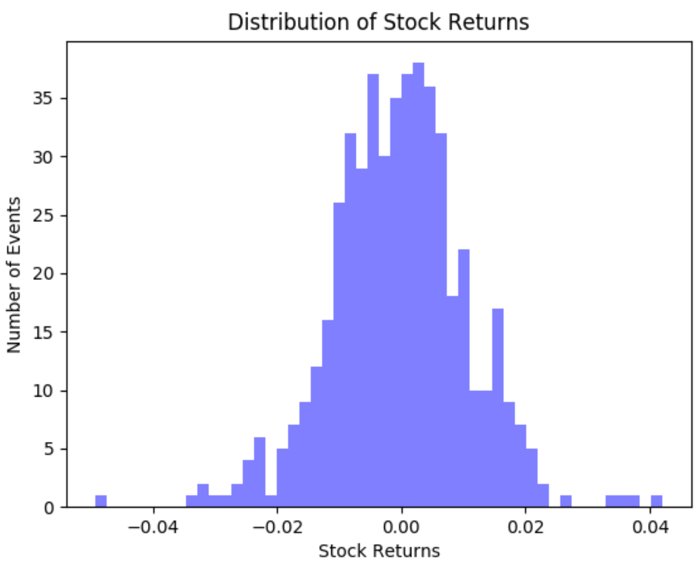

(b) AAPL




(c) GOOGL



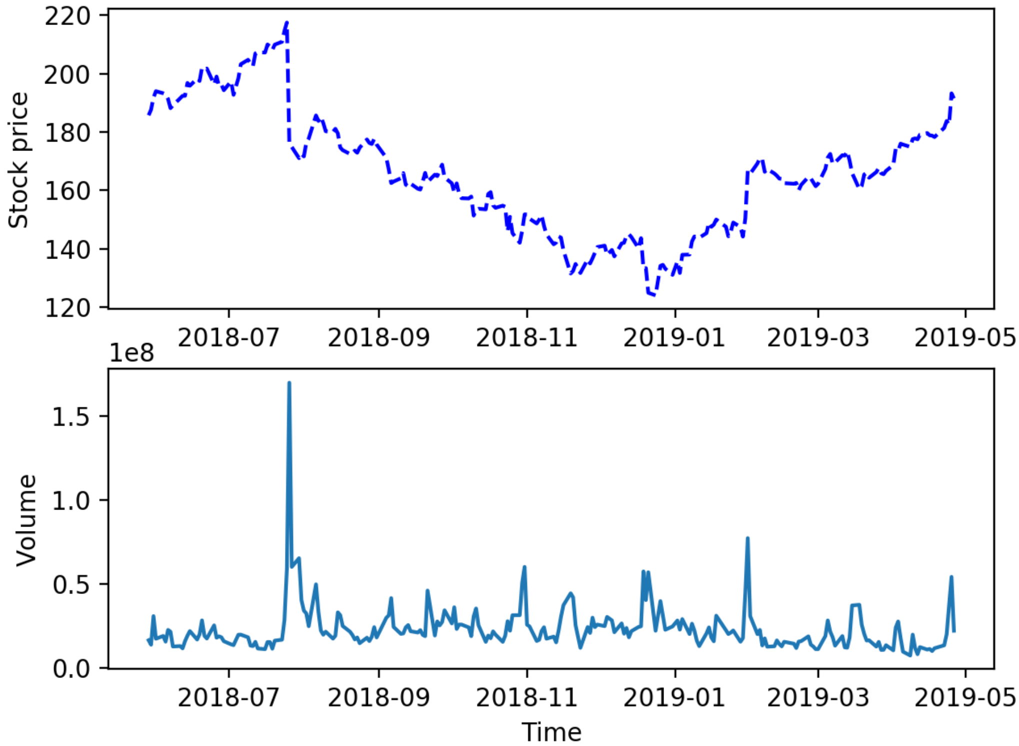

(d) AMZN



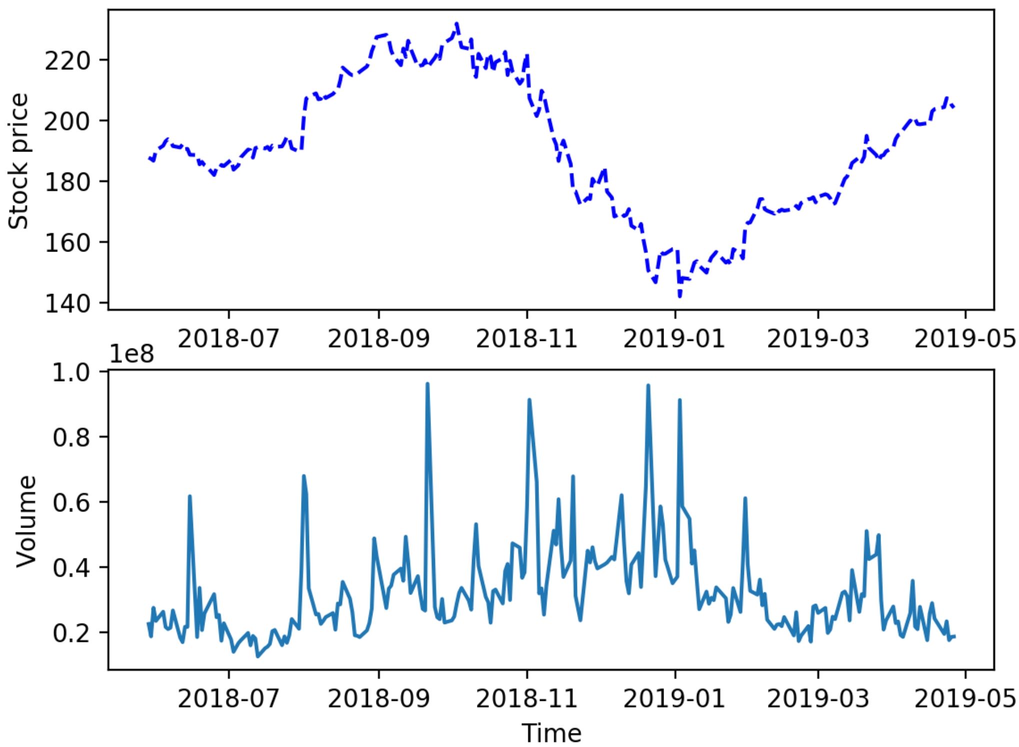

(e) AMJ



Figure 1: We choose FB, AAPL and GOOGL stock as examples to present the stock returns distribution. These three stock returns have asymmetric Lapalce distributions.

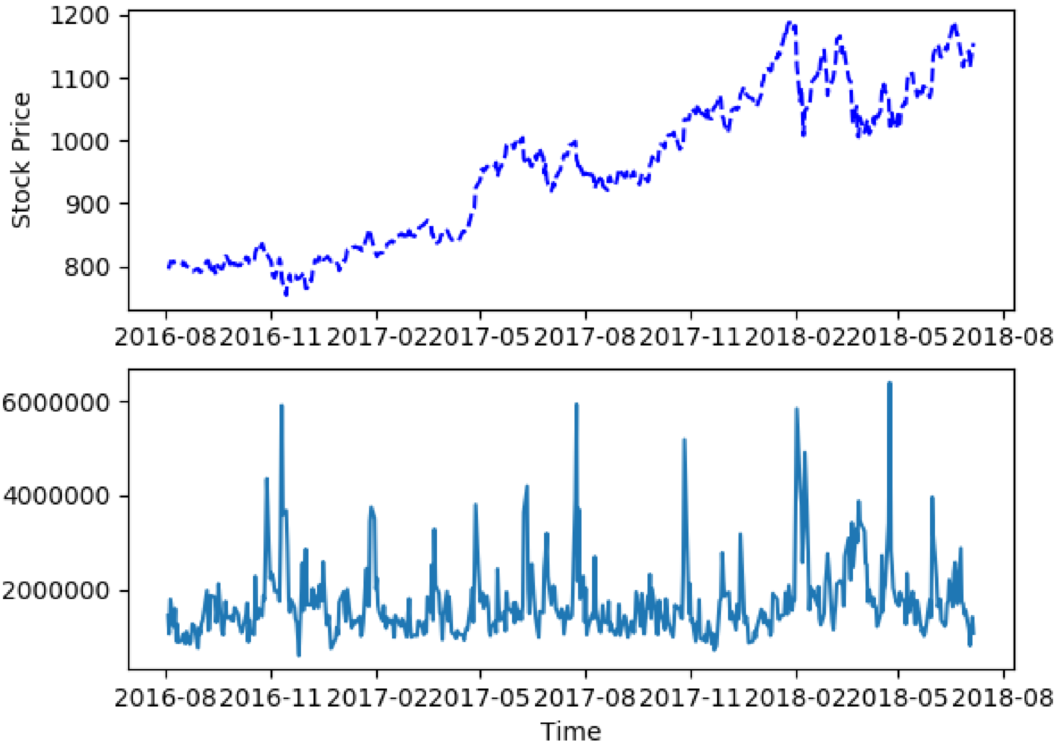



(a) FB

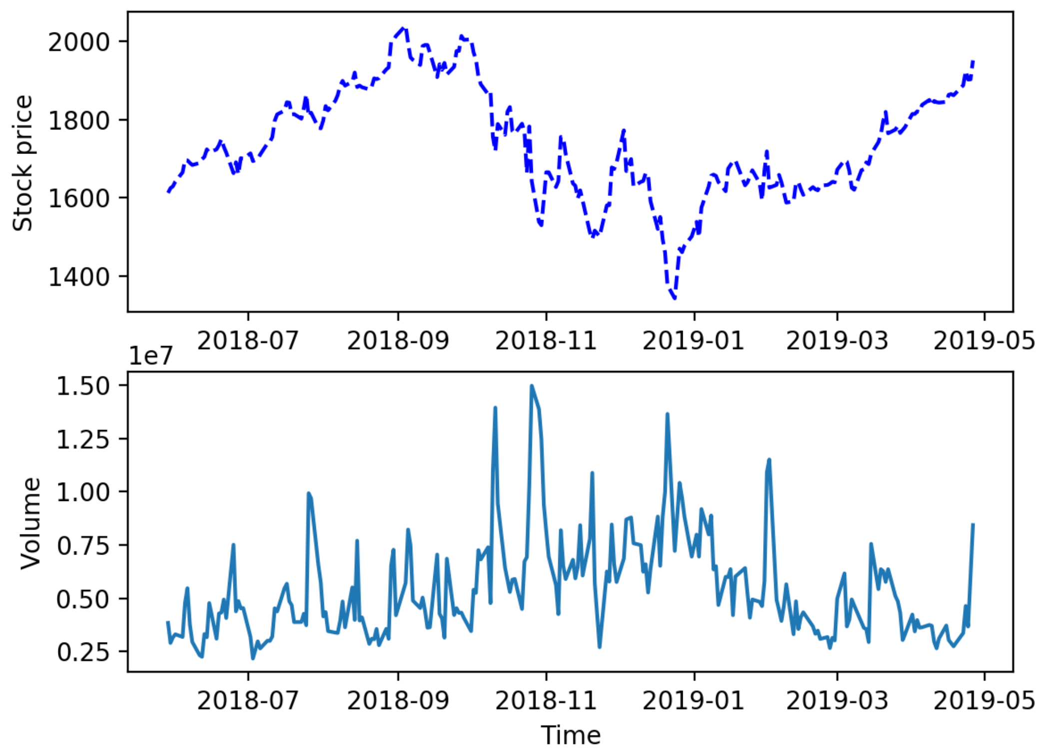



(b) AAPL

Figure 2: We selected (a) FB, (b) AAPL to demonstrate the relationship between the trading volume and the stock price. Most probable states for the two-regime ARHMM(1) model fitted on the (a) FB, (b) AAPL from 5/24/2018 to 5/24/2019, together with the fluctuation of trading volume. The reversion of stock price usually appear after the high trading volume.

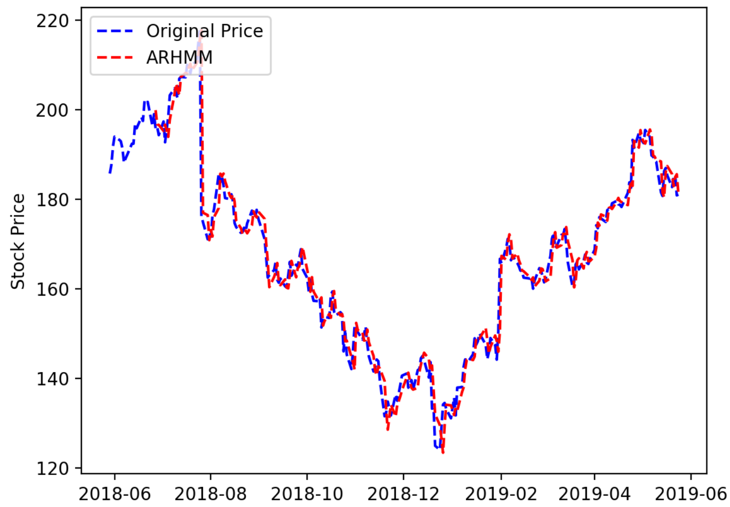



(a) GOOGL

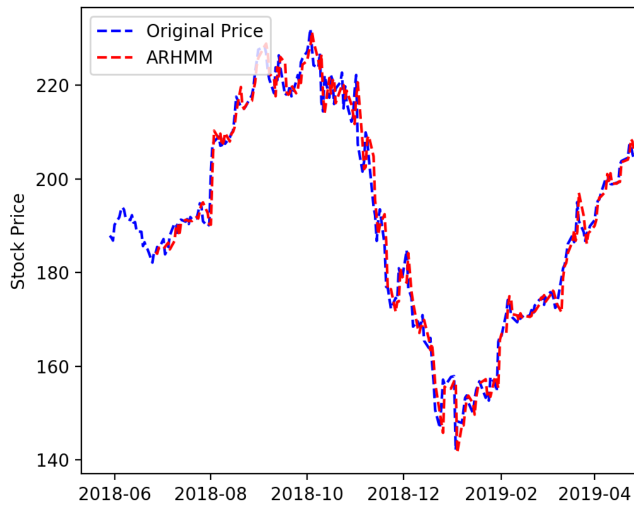



(b) AMZN

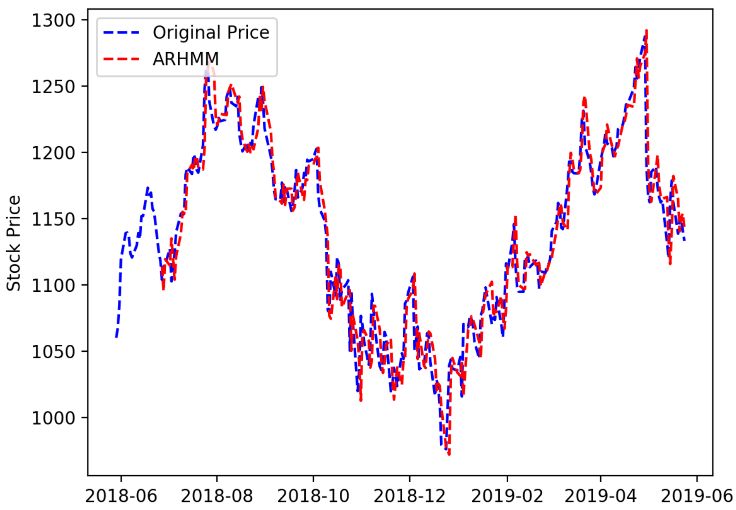
Figure 3: We selected (a) GOOGL, (b) AMZN to demonstrate the relationship between the trading volume and the stock price. Most probable states for the two-regime ARHMM(1) model fitted on the (a) GOOGL, (b) AMZN from 5/24/2018 to 5/24/2019, together with the fluctuation of trading volume. The reversion of stock price usually appear after the high trading volume.



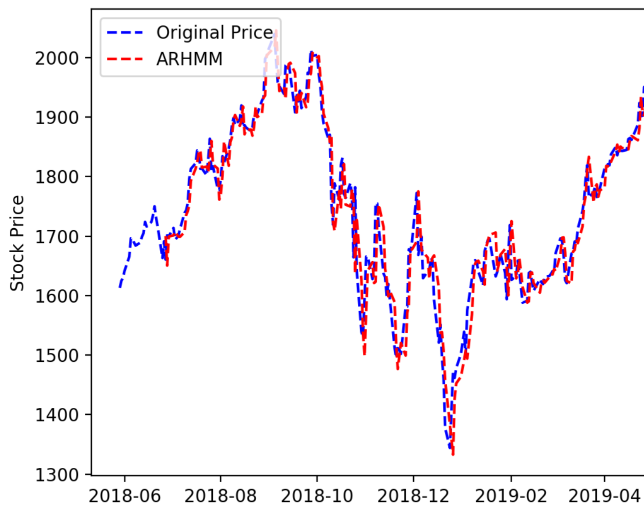

(a) FB



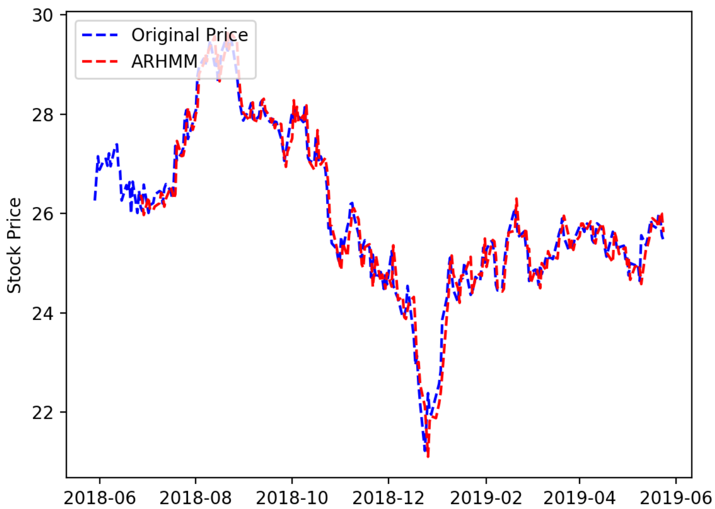

(b) AAPL




(c) GOOGL




(d) AMZN




(e) AMJ



Figure 4: We selected five stocks: , (a) FB, (b) AAPL, (c) GOOGL, (d) AMZN, and (e) AMJ, to show that our odel successfully tracks future stock prices.