2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC) | 978-1-6654-2463-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/COMPSAC51774.2021.00264

# A Novel Algorithmic Trading Strategy using Hidden Markov Model for Kalman Filtering Innovations

Ethan Johnson-Skinner
*Department of Mathematics*
*Ryerson University*
Toronto, Canada
ethan.johnsonskinner@ryerson.ca

You Liang
*Department of Mathematics*
*Ryerson University*
Toronto, Canada
you.liang@ryerson.ca

Na Yu
*Department of Mathematics*
*Ryerson University*
Toronto, Canada
nayu@ryerson.ca

Alin Morariu
*Department of Mathematics*
*Ryerson University*
Toronto, Canada
alin.morariu@ryerson.ca

***Abstract*—The development of algorithmic trading has been one of the most prominent trends in finance and its applications. Hidden Markov Models (HMMs) help enhance the predictive power of statistical models and improve trading strategies for data scientists and algorithmic traders. In recent years there has been growing interest in investigating the pairs trading and multiple trading based on robust Kalman filtering (KF) using data-driven innovation volatility forecasts (DDIVF). KF algorithms were successfully applied in pairs trading with two cointegrated assets using DDIVF as a method for forecasting non-normal innovation volatility. In this paper a novel combined pairwise trading strategy is proposed by combining HMM and DDIVF to further optimize trading signals in different market regimes. The results of the numerical experiments on two cointegrated stocks show that the proposed profitable trading strategy using DDIVF-HMM outperforms the recently studied robust trading strategy using DDIVF alone.**

***Index Terms*—Pairs Trading, Hidden Markov Models, Robust Kalman Filter, Innovation Volatility**

## I. Introduction

The stock market is complex and volatile. The price fluctuations depend on many factors, including equity, interest rate, inflation, treasure yields, options, earnings calendar, merger and acquisition of public listed companies etc. Therefore, building a model that accounts for many factors as possible is desirable. Algorithmic trading ( [3], [5], [9], [15]) not only incorporates the predictions of the further market and implements trading strategies to profit in various global markets, but also can produce profits at a speed and frequency that a human trader can not achieve. Pairs trading is used to exploit the securities that are out of equilibrium in financial markets. The strategy involves identifying two securities (e.g. stocks, bonds, foreign exchanges) whose prices tend to move together in the long term. When prices are divergent, the cheaper security is bought long and the more expensive one is sold short. When prices converge back to the equilibrium, the trade is ended and a profit is obtained.

Pairs trading was introduced to the academic community through [7] in 2006. The key idea behind pairs trading is closely linked with the statistical concept of cointegration. If a linear combination of a group of non-stationary time series is stationary, then the group is determined to be cointegrated. For cointegrated prices, $P_{1,t}$ and $P_{2,t}$, the difference or spread of two prices, $\epsilon_t = P_{1,t} - \beta_0 - \beta_1 P_{2,t}$, is stationary, which suggests that $\epsilon_t$ moves around an equilibrium value (about the the mean of $\epsilon_t$) . In pairs trading, the time varying regression coefficient $\beta_1$ is called the hedge ratio, and it describes the amount of one security to purchase or sell for every unit of the other security. The regression coefficients $\beta_0$ and $\beta_1$ are estimated with historical data along with an estimate of volatility, $\hat{\sigma}$, via the square root of mean square error (MSE). However, $\hat{\sigma}$ as a square root estimator is not efficient (see [14] for details) statistically. For a trading period, trading signals are generated by computing a z-score based on the regression coefficient estimates and volatility estimate, which is given by $z_t = (P_{1,t} - \hat{\beta}_0 - \hat{\beta}_1 P_{2,t})/\hat{\sigma}$.

In order to incorporate the time varying regression coefficients ( [2], [12]), and extend pairs trading to multiple trading [8], the linear state space model or dynamic linear model can be used. The state space model employs a random walk as the state equation:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{v}_t, \tag{1}$$

where $\boldsymbol{\beta}_t$ is the $m$-dimensional state vector at time $t$, and $\boldsymbol{v}_t$ is i.i.d with mean zero and covariance matrix $\Sigma_{\boldsymbol{v}}$. An observed process $y_t$, is described by an observation equation:

$$y_t = \boldsymbol{A}_t \boldsymbol{\beta}_t + \epsilon_t, \tag{2}$$

where $\boldsymbol{A}_t$ is a $m$-dimensional feature or predictor, and the observational noise $\epsilon_t$ is i.i.d with mean zero and variance $\sigma_\epsilon^2$. A primary purpose of the analysis is to derive dynamic filtered estimates, $\widehat{\boldsymbol{\beta}}_{t|t} = E[\boldsymbol{\beta}_t|\mathcal{F}_t^y]$, for the hedge ratio $\boldsymbol{\beta}_t$ to hedge the risk exposure of the stock price movement, given the data $\mathcal{F}_t^y = \{y_1, \ldots, y_t\}$ up to time $t$. Using the filtered estimate $\widehat{\boldsymbol{\beta}}_{t-1|t-1}$, $\nu_t = y_t - \boldsymbol{A}_t \widehat{\boldsymbol{\beta}}_{t-1|t-1}$ is called the innovation at time $t$. The innovation sequence $\nu_t$ and its time varying volatility are used to generate trading signals in algorithmic trading.

In recent years there has been growing interest in pairs trading and multiple trading based on Kalman filtering (KF) and maximum informative filtering. In the literature [2], [8], [12], very small initial values (e.g. $\sigma_\epsilon^2 = 0.001$) of the KF are used. However trading profit is sensitive to initial values, and it drops sharply when initial values increase slightly. [10] studied a novel resilient data-driven filtering algorithm based on regularized estimating functions for multiple trading, which does not require either very small initial values or

978-1-6654-2463-9/21/$31.00 ©2021 IEEE
DOI 10.1109/COMPSAC51774.2021.00264

the normality assumption for errors. [14] showed that the commonly used square root of the innovation variance is not an appropriate estimator of the innovation (non-normal) volatility. The data-driven generalized exponential weighted moving average (DD-EWMA) volatility forecasting model proposed in [14] is used in [10] to propose the data-driven innovation volatility forecast (DDIVF), and improve the stability of the filtering algorithm. The DDIVF provides accurate dynamic one-step ahead forecasts of innovations which are used to generate the trading signals appropriately. Let the conditional variance of the innovation $\nu_t$, based on the past data up to time $t-1$, be $\sigma_t^2$. The DD-EWMA volatility forecasting model for innovations is given by

$$\hat{\sigma}_t = (1-\alpha)\,\hat{\sigma}_{t-1} + \alpha\frac{|\nu_{t-1}-\bar{\nu}|}{\hat{\rho}_\nu}, \quad 0<\alpha<1, \tag{3}$$

where $\alpha$ is the tuning parameter, and $\hat{\rho}_\nu$ is the estimated sign correlation of the innovation sequence, defined as $\mathrm{Corr}(\nu_t - \bar{\nu}, \mathrm{sgn}(\nu_t - \bar{\nu}))$. The optimal value of $\alpha$ is obtained by minimizing the one-step ahead forecast error sum of squares (FESS), and the estimated sign correlation $\hat{\rho}_\nu$ is used to identify the conditional $t$ distribution of $\nu_t$. It was demonstrated in [10] that the pairs trading strategy using the DDIVF outperforms that using Kalman filter innovation volatility forecast (KFIVF). In this paper, model (3) is used and extended to study the volatility forecasts of the innovations to incorporate the hidden states of innovations.

Hidden Markov models (HMMs) have been widely used in the areas such as speech recognition, DNA sequencing, electrical signal prediction and image processing. HHMs have also been applied in finance including stock price prediction and option pricing. [6] implements a regime switching HMM approach for airline stock prices forecasting for interrelated markets. [11] discussed the use of HHMs to capture different regimes, and switch the model for option valuation based on each regime. Therefore, it is important to include regime switching in pairs trading and multiple trading. We combine the methodology of [11] and [10] to introduce a regime-aware pairwise DDIVF-HMM trading strategy where the threshold for trading signals is conditional on the hidden state of the KF innovations $\nu$.

The remainder of this paper is organized as follows. In Section II, a KF algorithm is proposed with DDIVF. A data-driven multiple trading strategy, using filtered hedge ratios and DDIVF-HMM, is proposed in section II as well. In Section III, the results of the numerical experiments on two cointegrated stocks show that pairs trading strategies constructed using the combined DDIVF-HMM with two or three hidden innovation states outperform the trading strategies using the DDIVF alone. These two strategies are analyzed and compared using a training sample and a test sample. Finally, Section IV provides conclusions.

## II. Methods

In this section, the multiple trading strategy using multiple cointegrated stocks is investigated. Consider $m$ asset prices $P_{1,t}, P_{2,t}, \ldots, P_{m,t}$ with a cointegrated relationship. The state space model (1) - (2) is used where $\boldsymbol{\beta}_t = (\beta_{0,t}, \beta_{1,t}, \ldots, \beta_{m-1,t})'$, $y_t = P_{1,t}$ and $\boldsymbol{A}_t = (1, P_{2,t}, \ldots, P_{m,t})$. In addition, it is assumed for simplicity that $\boldsymbol{\beta}_0, \boldsymbol{v}_t$ and $\epsilon_t$ are uncorrelated.

### A. Kalman Filters Using DDIVF

For model (1) - (2), let $\widehat{\boldsymbol{\beta}}_{t-1|t-1} = \mathrm{E}[\boldsymbol{\beta}_{t-1}|\mathcal{F}^y_{t-1}]$ and $P_{t-1|t-1} = \mathrm{Var}(\boldsymbol{\beta}_{t-1} - \widehat{\boldsymbol{\beta}}_{t-1|t-1}|\mathcal{F}^y_{t-1})$. Kalman filters or the non-Gaussian maximum informative filters in [8], [10] yield the optimal estimate of $\boldsymbol{\beta}_t$ as

$$\widehat{\boldsymbol{\beta}}_t = \widehat{\boldsymbol{\beta}}_{t-1|t-1} + (P_{t-1|t-1} + \Sigma_{\mathbf{v}})\boldsymbol{A}_t' Q_t^{-1}(y_t - \boldsymbol{A}_t\widehat{\boldsymbol{\beta}}_{t-1|t-1}), \tag{4}$$

where the innovation variance is given by

$$Q_t = \mathrm{Var}(\nu_t|\mathcal{F}^y_{t-1}) = \mathbf{A}_t(P_{t-1|t-1} + \Sigma_{\mathbf{v}})\mathbf{A}_t' + \sigma^2_\epsilon.$$

In most of the applications including pairs trading and risk forecasting, the filtered estimate $\widehat{\boldsymbol{\theta}}_{t-1|t-1}$ for the state variable, the innovation, $\nu_t$, and the innovation volatility, $\sqrt{Q_t}$, are used. However, $\sqrt{Q_t}$ is not an appropriate estimate of the innovation volatility. Therefore, DD-EWMA volatility forecasting model (3) is used to obtain DDIVF, and Algorithm 1 illustrates the details of DDIVF calculation. Based on the past $k$ innovations, $\nu_{t-k}, \ldots, \nu_{t-1}$, estimated sign correlation, $\hat{\rho}_\nu$, and volatility estimate $|\nu_s - \bar{\nu}|/\hat{\rho}_\nu, s = t-k, \cdots, t-1$, are calculated. The smoothed value $S_s$ of the volatility estimate is calculated recursively. The optimal smoothing constant, $\alpha_{opt}$, is determined by minimizing the one-step ahead FESS. Using the optimal value $\alpha_{opt}$, the smoothed value, $S_s$, is calculated recursively. Finally, $S_{t-1}$ is computed, and used as the volatility forecast $\hat{\sigma}_t^{DD}$ for $\nu_t$.

**Algorithm 1** Dynamic DD-EWMA volatility forecasts of innovation

**Require:** Predicted errors $\nu_s, s = t-k, \cdots, t-1$
1: $\hat{\rho}_\nu \leftarrow \mathrm{Corr}(\nu_s - \bar{\nu}, \mathrm{sign}(\nu_s - \bar{\nu}))$
2: $V_s \leftarrow |\nu_s - \bar{\nu}|/\hat{\rho}_\nu$ {Compute estimated volatility}
3: $S_{t-k-1} \leftarrow \bar{V}_l$ {Initial volatility forecast using first $l$ observations}
4: $\alpha \leftarrow (0.01, 0.5)$ *by* $0.01${Set a range for $\alpha$}
5: $S_s \leftarrow \alpha * V_s + (1-\alpha) * S_{s-1}, s = t-k, \ldots, t-1$
6: $\alpha_{opt} \leftarrow \min_\alpha \sum_{s=t-k+l}^{t-1}(V_s - S_{s-1})^2$
7: $S_s \leftarrow \alpha_{opt} * V_t + (1-\alpha_{opt}) * S_{s-1}, s = t-k, \ldots, t-1$
8: $\hat{\sigma}_t^{DD} \leftarrow S_{t-1}$
9: **return** $\alpha_{opt}, \hat{\sigma}_t^{DD}$

Algorithm 2 explains how to compute the dynamic filtered hedge ratio $\widehat{\boldsymbol{\beta}}_{t|t}$ using KF and to incorporate DDIVF $\hat{\sigma}_t^{DD}$ as the innovation volatility forecast. The innovation $\nu_t$, and the standard deviation $\sqrt{Q_t}$ or the DDIVF $\hat{\sigma}_t^{DD}$ can be used to construct the signals for a trading strategy at each time $t$. In the literature [2], [8], [12], trading profits using $\sqrt{Q_t}$ are sensitive to initial values. Therefore, a robust multiple trading strategy using $\widehat{\boldsymbol{\beta}}_{t-1|t-1}$, $\nu_t$ and DDIVF $\hat{\sigma}_t^{DD}$ is proposed in [10] and compared with the multiple trading strategy using

$\sqrt{Q_t}$ to demonstrate the profitability and robustness of the DDIVF approach. The dynamic robust z-score $z_t$ using $\hat{\sigma}_t^{DD}$ is computed as

$$z_t = \nu_t / \hat{\sigma}_t^{DD}, \tag{5}$$

and the z-scores will be compared with a threshold value $p$ to generate trading signals.

**Algorithm 2** Dynamic filtered hedge ratios and innovation volatility forecasts

**Require:** Data: adjusted closing stock prices $P_{1,t}, P_{2,t}, \ldots, P_{m,t}, t = 1, \ldots, n$

Let $y_t = P_{1,t}, \boldsymbol{A}_t = (1, P_{2,t}, \ldots, P_{m,t})$

Initialization: initial state $\boldsymbol{\beta}_0$, initial error covariance matrix $P_{0|0} = \Sigma_0$, constant error covariance matrix $\Sigma_{\mathbf{v}}$, constant innovation variance $\sigma_\epsilon^2$

**for** $t \leftarrow 1, \ldots, n$ **do**

  Prediction: Based on data available at $t-1$:

  $\widehat{\boldsymbol{\beta}}_{t|t-1} \leftarrow \widehat{\boldsymbol{\beta}}_{t-1|t-1}; P_{t|t-1} \leftarrow P_{t-1|t-1} + \Sigma_{\mathbf{v}}; \hat{y}_{t|t-1} \leftarrow \boldsymbol{A}_t \widehat{\boldsymbol{\beta}}_{t|t-1}$

  Update: Inference about $\boldsymbol{\beta}_t$ is updated using the observation $y_t$ at time $t$

  $\nu_t \leftarrow y_t - \hat{y}_{t|t-1}; Q_t \leftarrow \boldsymbol{A}_t P_{t|t-1} \boldsymbol{A}_t' + \sigma_\epsilon^2$

  DDIVF $\hat{\sigma}_t^{DD}$ is calculated based on $\nu_{t-k}, \ldots, \nu_{t-1}$ using Algorithm 1

  $\widehat{\boldsymbol{\beta}}_{t|t} \leftarrow \widehat{\boldsymbol{\beta}}_{t|t-1} + P_{t|t-1} \boldsymbol{A}_t' Q_t^{-1} \nu_t; P_{t|t} = (I - P_{t|t-1} \boldsymbol{A}_t' Q_t^{-1} \boldsymbol{A}_t) P_{t|t-1}$

**end for**

**return** $\widehat{\boldsymbol{\beta}}_{t|t}, \nu_t, \hat{\sigma}_t^{DD}$

### *B. HMM for Innovation*

Define $S_t$ as the hidden innovation state variable with $K$ possible values. Here we discuss how to apply a HMM to determine the value of the hidden state variable $S_t$ for each observed innovation $\nu_t$ at time $t$, $t = 1, 2, \ldots, n$. Under HMM, future hidden states depend only on the current hidden state by the Markov assumption. At each time $t$, an observation of $\nu_t$ is estimated and its corresponding state $S_t$ is determined depending on the previous state $S_{t-1}$, state transition probability $P(S_t|S_{t-1})$ and emission probability $P(\nu_t|S_t)$. The emission probabilities link the states (unobserved) to $\nu_t$ (observed).

To fit a HMM, the transition matrix of the Markov chain and emission probabilities must be estimated. Define the transition probability matrix of hidden states as

$$P = \{p_{ij}\} \text{ s.t. } p_{ij} = P(S_{t+1} = j | S_t = i) \tag{6}$$

where $i, j \in \{1, 2, \ldots, K\}$. Observable data $\nu_t$ is linked to the hidden state $S_t$ by emission probabilities $P(\nu_t|S_t)$. The joint density of the hidden states and observable data is given as

$$P(\nu_{1:n}, S_{1:n}) = P(S_1) \prod_{t=2}^{n} P(S_t|S_{t-1}) \prod_{t=1}^{n} P(\nu_t|S_t) \tag{7}$$

The expectation-maximization (EM) algorithm is used to estimate the transition matrix and emission probability parameters (see [1], [13], [16]). Define $\gamma_{i,t} = P(S_t = i|\nu_{1:t})$ as the probability of being in state $1, \ldots, K$ at time t. Let $\gamma_{il,t} = P(S_t = i, \nu_{it} = l|\nu_{1:t})$ where $\nu_{il}$ is a random variable indicating the mixture component at time $t$ for state $i$. The parameters $\mu$ and $\Sigma$ for the emission probabilities are obtained using the update equations: $\mu_{il} = \frac{\sum_{t=1}^{n} \gamma_{il,t} \nu_t}{\sum_{t=1}^{n} \gamma_{il,t}}$; $\Sigma_{il} = \frac{\sum_{t=1}^{n} \gamma_{il,t} (\nu_t - \mu_{il})(\nu_t - \mu_{il})^T}{\sum_{t=1}^{n} \gamma_{il,t}}$.

### *C. A Novel Multiple Trading Strategy using DDIVF-HMM*

This paper's novelty is highlighted in this subsection. The DDIVF approach presented in [10] is extended by incorporating a HMM to generate a dynamic fitting of the trading threshold ($p$). With the DDIVF approach the upper and lower trading bands are fitted with a single optimal value $p_{opt}$. DDIVF-HMM determines an optimal threshold value $p_{opt}^{HMM}$ for each hidden innovation state, $S_t = j, j \in \{1, \ldots, K\}, t = 1, \ldots, n$, that is used to build the trading signals. $p_{opt}^{HMM}[S_t]$ is defined as $p_{opt}^{HMM} \in \mathbb{R}^K$ and $p_{opt}|S_t = j, j \in \{1, \ldots, K\}$. The upper and lower trading bands are calculated using $p_{opt}^{HMM}[S_t]\hat{\sigma}_t^{DD}$, where $p_{opt}^{HMM}[S_t]$ is determined from the training data as explained in Algorithm 3. A flow chart in Fig. 1 characterizes the relationship between the KF hedge ratio, DDIVF $\hat{\sigma}_t^{DD}$, hidden innovations states, and the trading signal $z_t$.

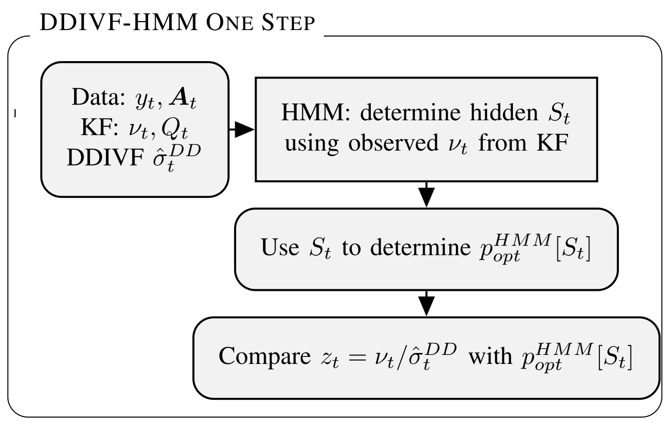


Fig. 1. The flow chart demonstrates the algorithm that integrates Kalman Filter, data driven innovation volatility forecast and hidden Markov model (KF-DDIVF-HMM) trading strategy.

KF, DDIVF and the HMM are integrated to create a novel combined trading strategy, DDIVF-HMM as shown in Fig. 1. The key idea of the combined DDIVF-HMM approach is that the HMM at time $t$ is built based on the innovation $\nu_t$ and $p_{opt}^{HMM}[S_t]$ is used to generate the trading signal at time $t$. The DDIVF-HMM trading strategy is further explained in Algorithm 3. Sells are represented as $s_t = -1$, buys as $s_t = 1$, and no signal as $s_t = 0$. Sell signal ($s_t = -1$, sell more expensive $P_{1,t}$) is generated when $z_t$ crosses $p^{HMM}[S_t]$ from below, or equivalently, $\nu_t$ crosses $p^{HMM}[S_t]\hat{\sigma}_t^{DD}$ from below. Buy signal ($s_t = 1$, buy cheaper $P_{1,t}$) is generated when $z_t$ crosses a threshold $-p^{HMM}[S_t]$ from above, or equivalently, $\nu_t$ crosses $-p^{HMM}[S_t]\hat{\sigma}_t^{DD}$ from above. Trading positions are determined using $s_t$, and profit is computed. The $p$ vector is from 0.5 to 2.5 with increments by 0.1. A brute force method is used when finding the optimal $p_{opt}^{HMM}$, where the combination of the values in the vector $p$ are created in a matrix with $P^K$ being the number of rows and $K$ being the number of columns. We use the notation $p_{P^K,K}^{HMM}$ to denote the matrix of $p$ for training the DDIVF-HMM approach.

**Algorithm 3** DDIVF-HMM trading strategy

**Require:** $p_{P^K,K}^{HMM}$, $P_{1,t},\dots,P_{m,t}, t=1,\dots,n$
Compute $\nu_t$, $\hat{\sigma}_t^{DD}$ with Algorithm 2
Determine $S_t$ with input $\nu_t$
**for** length of $P^{HMM}$ **do**
  Generate trading signals $s_t$:
  **for** $t \leftarrow k+2,\dots,n$ **do**
    If $\nu_t > p^{HMM}[S_t]\hat{\sigma}_t^{DD}$ & $\nu_{t-1} < p^{HMM}[S_{t-1}]\hat{\sigma}_{t-1}^{DD}$, then $s_t \leftarrow -1$
    If $\nu_t < -p^{HMM}[S_t]\hat{\sigma}_t^{DD}$ & $\nu_{t-1} > -p^{HMM}[S_{t-1}]\hat{\sigma}_{t-1}^{DD}$, then $s_t \leftarrow 1$
    Else $s_t \leftarrow 0$
    $position.A_t \leftarrow -1000 * \widehat{\boldsymbol{\beta}}_{t-1|t-1} * s_t$; $position.y_t \leftarrow 1000 * s_t$
    $profit.A_t \leftarrow (\boldsymbol{A}_t - \boldsymbol{A}_{t-1}) * position.A_t$; $profit.y_t \leftarrow position.y_t * (y_t - y_{t-1})$
    $profit_t \leftarrow profit.A_t + profit.y_t$
  **end for**
  Calculate the ASR as $ASR(P^{HMM}) = \sqrt{252} * mean(profit_t)/sd(profit_t)$
**end for**
Determine the optimal value $p_{opt}^{HMM}$ that maximizes $ASR$
Obtain the cumulative profit using $p_{opt}^{HMM}$
**return** $p_{opt}^{HMM}$, cumulative profit

## III. Results

This section compares the DDIVF-HMM pairs trading strategy and the DDIVF pairs trading strategy. The proposed methods and algorithms are illustrated using the adjusted closing prices of two cointegrated stocks downloaded from **Yahoo Finance** for the period from 2018-01-01 to 2021-01-01: Duke A. O. Smith Corp (AOS) and Energy Corp (DUK) as shown in Fig. 2. The training data is selected from 2018-01-01 to 2020-02-09 and the test data is selected from 2020-02-10 to 2021-01-01. The cointegration of the two stocks are regularly checked by the Engle-Granger test and the Johansen test over time. The training data is used to obtain $p_{opt}^{HMM}$ for each hidden state of the innovation. Then the test data is used to test the profitability and robustness of the proposed DDIVF-HMM trading strategy. It is known that March 2020 was a historically volatile month for the stock market. The proposed strategy is demonstrated to be profitable and robust during this period. The vertical blue line in 2 is used to show the separation between the train and test data set.

### A. DDIVF Pairs Trading Strategy

The DDIVF strategy is performed as follows. First, a rolling window approach is applied to the training data to forecast the volatility of $\nu_t$ using Algorithm 1. The selected data covers 531

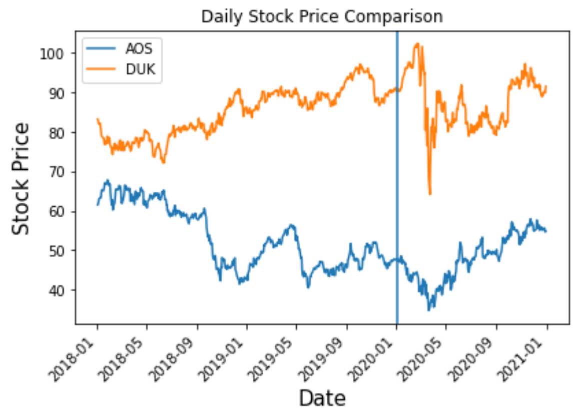


Fig. 2. Daily adjusted closing prices of AOS and DUK. The vertical straight line separates the training data and test data.

trading days, with 431 overlapping rolling windows, where each rolling window is 100 days and is used to calculate a one-day-ahead DDIVF. For example, $\nu_1,\dots,\nu_{100}$ are used to calculate the volatility forecast $\hat{\sigma}_{101}$ for $\nu_{101}$. The range of the tuning parameter $\alpha$ is chosen as $(0.1, 2)$ with an increment of 0.01. The single optimal threshold value is determined as $p_{opt} = 0.83$ using DDIVF. The corresponding optimal trading signals are visualized in Fig. 3, with upper trading band in orange and the lower trading band in green. The upper band is calculated using $\pm p_{opt}\hat{\sigma}_t^{DD}$ and the lower band is calculated using $\pm p_{opt} - \hat{\sigma}_t^{DD}$. Trading signals (i.e. sell when $s_t = -1$ and buy when $s_t = 1$) are generated when innovation $\nu_t$ crosses the upper band from bottom, or crosses the lower band from above. The look-ahead bias is eliminated by lagging the signals. Each trade consists of 1,000 units of the spread.

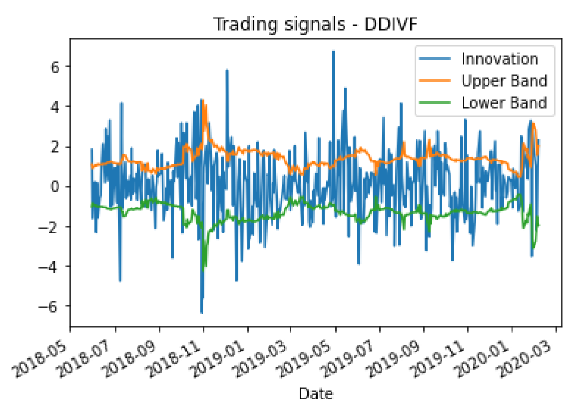


Fig. 3. Trading signals, $s_t$, are obtained using DDIVF and training data from 2018-01-01 to 2020-02-09.

With the training data, the optimal threshold value is $p_{opt} = 0.83$ as shown in Table I. The cumulative profit of the robust pairs trading strategy using DDIVF during the training period is $31722, which is higher than a buy and hold (B/H) strategy with a profit of $-1306 over the same time span.

TABLE I
TRADING PERFORMANCE: 2018-01-01 TO 2020-02-09 (TRAINING DATA)

| | $p_{opt}$ | ASR | Profit | B/H Profit |
|---|---|---|---|---|
| AOS/DUK | 0.83 | 3.1 | 31722 | -1306 |

$p_{opt}$: the optimal threshold value using DDIVF
ASR: annualized Sharpe ratio; B/H: buy and hold

In Table II, using $p_{opt} = 0.83$ from the training data, the trading signals are visualized in Fig. 4 for the test period. The profit and ASR of the DDIVF strategy and B/H strategy are compared for the test period in Table II. The cumulative profit of the DDIVF pairs trading strategy during the training period is \$5011, which is lower than B/H with a profit of \$24319.

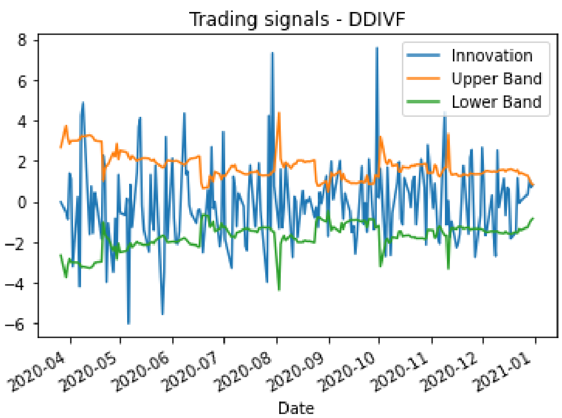


Fig. 4. Trading signals, $s_t$, are obtained using DDIVF and test data from 2020-02-10 to 2021-01-01.

TABLE II
TRADING PERFORMANCE: 2020-02-10 TO 2021-01-01 (TEST DATA)

| | $p_{opt}$ | ASR | Profit | B/H Profit |
|---|---|---|---|---|
| AOS/DUK | 0.83 | 0.86 | 5011 | 24319 |

### *B. DDIVF-HMM Pairs Trading Strategy with Two Hidden Innovation States*

The proposed DDIVF-HMM pairs strategy is validated with the same adjusted closing prices of AOS and DUK. The HMM state $S_t$ at time $t$ is found from the innovation estimate $\nu_t$. The training data is used to obtain $p_{opt}^{HMM}$ for each of the HMM states. We begin with a two state HMM before expanding the scope to three. The number of states of the model is a hyper-parameter that can be tuned with the training data and chosen based on the maximum annual sharp ratio.

We first explore two hidden states of the innovations, and adjust the upper and lower trading bands. Both HMM state 0 and state 1 (yellow and red trajectories) have their respective values of $p_{opt}^{HMM}$ that are fitted from the training data. The orange lower and blue upper trading bands are modelled from $\pm p_{opt}^{HMM}[S_t]\hat{\sigma}_t^{DD}$ in Fig. 5. Trading signals -1/1 are generated when innovation $\nu_t$ crosses the upper band from bottom, or crosses the lower band from above. The optimal threshold values using the DDIVF-HMM method are $p_{opt}^{HMM}$ = {state 0: 0.5, state 1: 1.7} as in Table III. The cumulative profit using DDIVF-HMM during the training period is \$50338, higher than B/H with a profit of \$-1306.

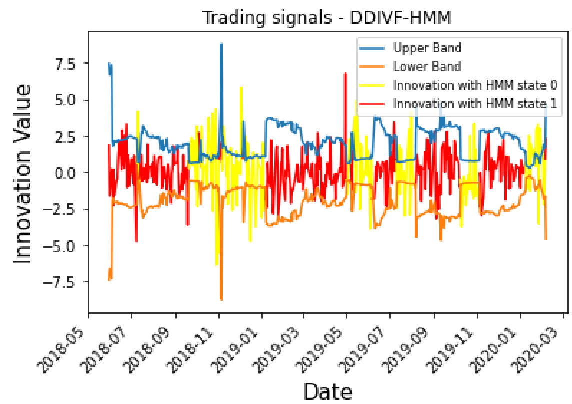


Fig. 5. Trading signals, $s_t$, are obtained from DDIVF-HMM and trained on: 2018-01-01 to 2020-02-09. Upper and lower bands are adjusted based on the optimal $p_{opt}^{HMM}$ for each state $S_t$ at time $t$.

TABLE III
TRADING PERFORMANCE: 2018-01-01 TO 2020-02-09 (TRAINING DATA)

| | $p_{opt}^{HMM}$ | ASR | Profit | B/H Profit |
|---|---|---|---|---|
| AOS/DUK | 0.5, 1.7 | 1.15 | 50338 | -1306 |

$p_{opt}^{HMM}$: optimal threshold value for each hidden innovation state

Fig. 6 visualizes the trading signals using the DDIVF-HMM trading strategy during the test period using $p_{opt}^{HMM}$ = {state 0: 0.5, state 1: 1.7} determined from the training data. Both HMM state 0 and state 1 (yellow and red trajectories) innovations have their respective values of $p_{opt}^{HMM}$ that are determined from the training data. The upper and lower trading signals have been adjusted by the respective value of $p_{opt}^{HMM}$ given the certain hidden state of the innovations.

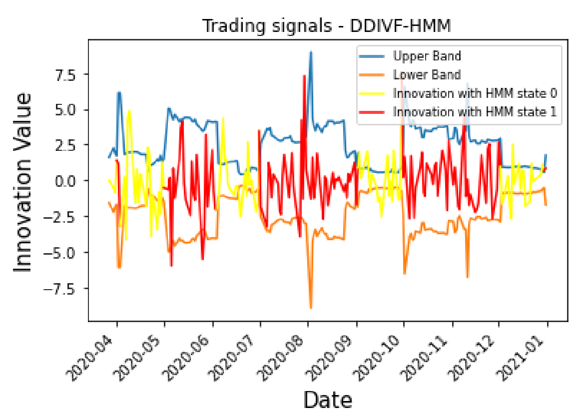


Fig. 6. Trading signals, $s_t$, are obtained using DDIVF-HMM and test data from 2020-02-10 to 2021-01-01.

As shown in Table IV, the cumulative profit using DDIVF-HMM during the test period is \$33830, higher than B/H strategy with a profit of \$24314. Comparing to profit of the DDIVF trading strategy in Table II, DDIVF-HMM outperforms DDIVF because the DDIVF-HMM trading strategy is able to capture the time varying innovation volatility as well as innovation state switching. Another interesting conclusion is that of the values of $p_{opt}^{HMM}$ change when the hidden states of the innovations change from volatile to less volatile market states. $p_{opt}^{HMM} = 1.7$ implies that less trades are made when market conditions are more volatile (highlighted in red in Fig. 5 and Fig. 6). During the less volatile period (highlighted in yellow in Fig. 5 and Fig. 6), $p_{opt}^{HMM} = 0.5$ allows for more trades to be made. It is consistent with human intuition that one normally makes more trades during a non-volatile period and less trades during a volatile period.

TABLE IV
TRADING PERFORMANCE: 2020-02-10 TO 2021-01-01 (TEST DATA)

| | $p_{opt}^{HMM}$ | ASR | Profit | B/H Profit |
|---|---|---|---|---|
| AOS/DUK | 0.5, 1.7 | 1.45 | 33830 | 24319 |

### C. DDIVF-HMM Paris Trading Strategy with Three Hidden States

Similarly, the DDIVF-HMM pairs trading strategy with three hidden states of innovations is discussed. It is determined from the training data that the optimal threshold value $p_{opt}^{HMM}$ = {state 0: 0.9, state 1: 2.1, state 2: 0.5}, which are used to adjust the upper and lower trading bands by $p_{opt}^{HMM}[S_t]\sigma_t^{DD}$. For the training period, the cumulative profit of the robust pairs trading strategy using DDIVF-HMM with three hidden states during the training period is \$54449 shown in Table V, higher than \$50338 using DDIVF-HMM with two hidden states shown in Table III.

TABLE V
TRADING PERFORMANCE: 2018-01-01 TO 2020-02-07 (TRAINING DATA)

| | $p_{opt}^{HMM}$ | ASR | Profit | B/H Profit |
|---|---|---|---|---|
| AOS/DUK | 0.9, 2.1, 0.5 | 1.22 | 54449 | -1306 |

For test period, the cumulative profit of the trading strategy using DDIVF-HMM with three hidden states during the training period is \$57621, higher than B/H with a profit of \$24319 shown in Table VI. The profit of \$57621 is also higher than that of \$33830 using DDIVF-HMM with two hidden states shown in Table IV. More than three hidden states will not be discussed because a larger number of hidden states will cause the over-fitting problem. The results of two and three hidden innovation states show that both two and three hidden Markov states for the pairs trading algorithm is more profitable than B/H strategy, and DDIVF-HMM outperforms DDIVF.

## IV. CONCLUSION

This paper presents a robust pairs trading strategy using KF as well as DDIVF based on DD-EWMA forecasts for each hidden state of the proposed HMM of the innovation. The driving idea, unlike the existing work in [10], is combining the DDIVF and HMM for innovations. We demonstrate that the DDIVF-HMM trading strategy is able to capture the time varying innovation volatility as well as innovation state switching. Numerical experiments are conducted to show that the proposed DDIVF-HMM trading strategy outperforms (i.e., more profitable) the DDIVF trading strategy.

TABLE VI
TRADING PERFORMANCE: 2020-02-08 TO 2021-01-01 (TEST DATA)

| | $p_{opt}^{HMM}$ | ASR | Profit | B/H Profit |
|---|---|---|---|---|
| AOS/DUK | 0.9, 2.1, 0.5 | 2.48 | 57621 | 24319 |

## REFERENCES


[1] J. Bulla, Hidden Markov models with t components. Increased persistence and other aspects, Quantitative Finance, vol. 11, (3), pp. 459-475, 2011.

[2] E. Chan, Algorithmic Trading: Winning Strategies and Their Rationale. Hoboken, New Jersey, John Wiley and Sons, 2013.

[3] C. Conlan, Automated Trading with R: Quantitative Research and Platform Development. Apress, 2016.

[4] J. Danielsson, K. R. James, M. Valenzuela and I. Zer, "Model risk of risk models", Journal of Financial Stability, vol. 23, pp. 79-91, 2016.

[5] M. L. Halls-Moore, Advanced Algorithmic Trading. Available at https://www.quantstart.com/advanced-algorithmic-trading-ebook/, 2017.

[6] Md. R. Hassan and B. Nath, "Stock Market Forecasting Using Hidden Markov Model: A New Approach", In proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications (ISDA05), 2005.

[7] E. G. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, "Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies", vol. 19(3), pp. 797-827, 2006.

[8] Y. Liang, A. Thavaneswaran, N. Yu, M. E. Hoque, and R. K. Thulasiram, "Dynamic data science applications in optimal profit algorithmic trading", pp. 1294-1299, In proceedings (workshop) of IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 2020.

[9] Y. Liang, A. Thavaneswaran, A. Paseka, Z. Zhu, and R. K. Thulasiram, "A novel data-driven algorithmic trading strategy using joint forecasts of volatility and stock price", pp. 293-302, In proceedings (Symposia) of IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 2020.

[10] Y. Liang, A. Thavaneswaran, and M. E. Hoque, "A Novel Algorithmic Trading Strategy Using Data-Driven Innovation Volatility", In proceedings of The 2020 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2020), pp. 1107-1114, 2020.

[11] C. C. Liew, T. K. Siu, "A hidden Markov regime-switching model for option valuation", Insurance: Mathematics and Economics, vol. 47(3), pp. 374-384, 2010.

[12] K. Longmore, Kalman Filter Example: Pairs Trading in R. Robot Wealth. https://robotwealth.com/kalman-filter-pairs-trading-r/, September, 2019.

[13] A. Thavaneswaran, A. A. Appadoo and J. Singh, "Random Coefficient Mixture (RCM) GARCH Models", Mathematical and Computer Modelling, vol 42, pp. 519-532, 2005.

[14] A. Thavaneswaran, A. Paseka and J. Frank, "Generalized Value at Risk Forecasting", Communications in Statistics Theory and Methods, pp. 1-8, 2019.

[15] A. Thavaneswaran, Y. Liang, Z. Zhu and R. K. Thulasiram, "Novel Data Driven Fuzzy Algorithmic Volatility Forecasting Models with Applications to Algorithmic Trading", In the proceeding of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, July, 2020.

[16] G. Xuan, W. Zhang and P. Chai, "EM algorithms of Gaussian mixture model and hidden Markov model," Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205), vol.1, pp. 145-148, 2001.