

# Pairs Trading with Robust Kalman Filter and Hidden Markov Model: Reproduction and Refinement of the Strategy of Johnson-Skinner et al.

Kaichong Wang

*Supervised by*

Prof. Daniel Totoum-Tangho

March 27, 2023

## 1 Overview

Statistical arbitrage strategies, such as pairs trading, have gained popularity in recent years. In a recent study, Johnson-Skinner, E., Liang, Y., Yu, N., & Morariu, A. (2021, July) [1] proposed a novel algorithmic trading strategy that applies a robust Kalman filter (KF) using data-driven innovation volatility forecasts (DDIVF) to forecast the hedge ratio and the volatility of the spread simultaneously. The strategy also uses the hidden Markov model (HMM) to optimize the thresholds for trading signals in different market regimes. This project aims to replicate and refine this strategy, using more rigorous mathematical reasoning and extensive trading simulations. In addition, we extended the empirical backtesting to twelve cointegrated pairs. My findings suggest that the strategy is not effective in real-world trading scenarios.

The rest of this report is organized as follows: Section 2 provides an introduction to the methodology of the proposed strategy, including a detailed explanation of the KF and HMM models, as well as the DDIVF approach. In Section 3, I present the results of our reproduction of the strategy and provide a comprehensive analysis of its performance, including statistics including annualized Sharpe Ratio, hit rate and number of trades, average win and loss, and annualized average return. Finally, in Section 4, I discuss the reasons why the strategy failed to produce satisfactory results and propose possible improvements for future research.

## 2 Methodology

This section is largely referenced from the paper [1] and [2]. To provide a more comprehensive understanding of the cointegration pairs trade and KF algorithms, I have unfolded some details in the methodology. Additionally, I have made some refinements to the original mathematical reasoning and the trading simulation. The methodology includes five main parts: (1) cointegration pairs trade and pairs selection, (2) KF and hedge ratio estimate, (3) DDIVF, (4) HMM and regime-aware thresholds, and (5) trading strategy and simulation. In the following subsections, I will describe each part of the methodology in detail.

### 2.1 Cointegration Pairs Trade and Pairs Selection

The key idea behind cointegration pairs trading is to apply a mean-reverting strategy to a stationary spread formed as a portfolio of a cointegrated pair of stocks with non-stationary prices. The Engle-Granger two-step approach [3] is used to identify the cointegrated pairs and determine the positions of each stock in the portfolio. This approach involves regressing the price series of one stock on the other using OLS and testing the stationarity of the residual series using the ADF test. To improve the accuracy of this process, I suggest using the log of prices instead of the original prices since the log of prices follows a geometric Brownian motion and is normally distributed, which better fits the assumptions of OLS.

For prices of a cointegrated stock pair,  $P_{1,t}$  and  $P_{2,t}$ , the spread  $\epsilon_t = \ln P_{2,t} - \beta_0 - \beta_1 \ln P_{1,t}$  is stationary, indicating that its value moves around an equilibrium value, and a mean-reverting strategy on the spread can generate profits from temporary divergence from this equilibrium. The “cointegration regression” coefficient  $\beta_1$ , also known as the hedge ratio, can be used to engineer the positions in the portfolio. Specifically, a long position in stock 2 and a short position of hedge ratio times the investment in stock 1 creates the desired exposure to the spread.

Trading signals are generated by computing a z-score based on the spread  $\epsilon_t$  and the estimated volatility of the spread  $\hat{\sigma}_t$ . Specifically,  $z_t = \epsilon_t / \hat{\sigma}_t$ . The cointegrated stock pairs are selected based on a reasonably small p-value (e.g., 10%) obtained from the Engle-Granger two-step approach.

### 2.2 Kalman Filter and Hedge Ratio Estimate

The cointegration relationship between the two assets may vary over time. To account for this time-varying relationship, a popular approach is to apply a linear state space model to the “cointegration regression” coefficients, as done in previous studies (e.g., [4] and [5]). The paper assumes that the “cointegration regression” coefficients follow a random walk process:

$$\beta_t = \beta_{t-1} + w_t$$

Using this assumption, we can estimate the hedge ratios using the Kalman Filter algorithm. The following description and notation of the Kalman Filter are referenced and adapted from KalmanFilter.Net [6].

The Kalman Filter is an algorithm that recursively updates an estimate of the state of a dynamic system based on prior knowledge from intermittent noisy measurement. In our case, the state corresponds to the “cointegration regression” coefficients and the measurements correspond to the observed prices of the pair of stocks. The Kalman Filter operates in a “predict-correct” loop. Once initialized, it predicts the system state and the uncertainty at the next step based on the state extrapolation equation and covariance extrapolation equation. When a measurement is received, the Kalman Filter updates (or corrects) the prediction and the uncertainty of the current state based on the measurement equation, the state update equation, and the covariance update equation. The Kalman Filter then predicts the following states and so on. The five equations that describe this process are as follows:

- State extrapolation equation:  $\hat{x}_{t+1|t} = \hat{x}_{t|t} + w_t$ , where  $\hat{x}$  is the predicted or estimated state vector, and  $w$  is the process noise vector.
- Covariance Extrapolation Equation:  $P_{t+1|t} = P_{t|t} + Q$ , where  $P$  is the predicted or estimated state uncertainty covariance matrix, and  $Q$  is the constant process noise matrix.
- Measurement Equation:  $z_t = H_t x_t + e_t$ , where  $z$  is the measurement vector,  $H$  is the observation matrix,  $x$  is the true system state (hidden state), and  $e$  is a random noise vector.
- State Update Equation:  $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t v_t$ , where  $K$  is the Kalman Gain and  $K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R)^{-1}$ , and  $R$  is the constant measurement uncertainty.  $v_t$  is the Kalman Filter innovation with  $v_t = z_t - H_t \hat{x}_{t|t-1}$ .
- Covariance Update Equation:  $P_{t|t} = (I - K_t H_t) P_{t|t-1} (I - K_t H_t)^T + K_t$ , where  $I$  is an identity matrix.

Our problem can be modeled using the Kalman Filter algorithm shown above by setting  $\hat{x}_{t|t-1} = [\hat{\beta}_0 \ \hat{\beta}_1]^T$ ,  $z_t = [\ln(P_{2,t})]$  and  $H_t = [1 \ \ln(P_{1,t})]$ . the Kalman Filter innovation, a scalar in this case,  $v_t$  represents the spread.

The measurement noise  $e_t$  represents the deviation of the actual observation from the true state at time  $t$ , which is not completely predictable. The measurement noise covariance matrix  $R$  captures the uncertainty of the measurement noise, which is usually assumed to be a constant value. In our problem, it is assumed that the measurement noise follows a Gaussian distribution with mean zero and variance  $\sigma_e^2$ . The covariance matrix  $R$  is therefore a one-by-one scalar matrix with the value  $\sigma_e^2$ . The process noise  $w_t$  represents the deviation of the predicted state from the true state at time

$t$ , which is due to unmodeled dynamics or random fluctuations in the system. The process noise covariance matrix  $Q$  captures the uncertainty of the process noise, which is usually assumed to be a diagonal matrix with non-zero values only along the diagonal. In our problem, we assume that the process noise follows a Gaussian distribution with mean zero and covariance matrix  $Q$ , which is a diagonal matrix with elements  $/ (1)$ . The paper arbitrarily assumed that  $= 0.0001$  and  $\sigma_e^2 = 0.001$ . However, these assumptions are not required, as they only serve to smooth the recursive estimation process and do not significantly affect the hedge ratio estimate  $\hat{x}_{t|t-1}$  or the Kalman Filter innovations  $v_t$ , as demonstrated in the numerical experiments in the next section.

To initialize the Kalman filter, the initial state estimate  $\hat{x}_{0|0}$  and the initial state uncertainty covariance matrix  $P_{0|0}$  need to be specified. In the original paper, both are set to zero. However, to improve the estimate, I used the “cointegration regression” coefficient estimate obtained from the first  $j = 100$  observations as the initial state estimate, since it is likely to be a better estimate than zero. The initial state uncertainty is still estimated as zero in this project. Further research may investigate better ways to initialize the estimated covariance of the two “cointegration regression” coefficients.

## 2.3 Data-driven Innovation Volatility Forecasting

The data-driven innovation volatility forecast (DDIVF) model is adapted from the data-driven generalized exponential weighted moving average (DD-EWMA) volatility forecasting model proposed in [7]. It provides a more robust volatility forecast than the conventional method. Conventionally, the volatility of the Kalman Filter innovations  $\sigma_t$  is estimated by  $\sqrt{H_t(P_{t-1|t-1} + Q)H_t^T + RR^T}$ . However, this estimate relies on the normality assumption of the innovation and accurate assumptions on process noise and measurement noise.

The DDIVF model identifies volatility by the relationship between absolute deviation and standard deviation for symmetric distributions with finite variance, and estimates the volatility as an exponential weighted moving average, assuming that the volatility is mean-reverting. Denote the conditional variance of the innovation  $v_t$ , based on the past data up to time  $t - 1$ , by  $\sigma_t$ . The volatility estimate at time  $t$  is given by the following equation:

$$\hat{\sigma}_t = (1 - \alpha)\hat{\sigma}_{t-1} + \alpha \frac{|v_{t-1} - \bar{v}|}{\hat{\rho}_v}$$

Here,  $\alpha$  is the smoothing constant ranging from zero to one, and  $\hat{\rho}_v$  is the sample sign correlation of the innovation sequence and  $\hat{\rho}_v = \text{corr}(v_t - \bar{v}, \text{sgn}(v_t - \bar{v}))$ . The sample sign correlation  $\hat{\rho}_v$  is used to identify the conditional distribution of  $v_t$ . The smoothing constant  $\alpha$  is obtained by minimizing the one-step ahead forecast error sum of squares (FESS).

The DDIVF algorithm uses the past  $k = 100$  innovations  $v_{t-k}, \dots, v_{t-1}$  to calculate the sample sign correlation  $\hat{\rho}_v$  and the volatility estimate  $|v_s - \bar{v}|/\hat{\rho}_v$ ,  $s = t - k, \dots, t - 1$ . The smoothed volatility estimate  $S_s$  is calculated recursively, with

the initial value being the average volatility estimates of the first  $l = 100$  observations. The optimal smoothing constant  $\alpha_{opt}$  is determined by minimizing the one-step ahead FESS. Using  $\alpha_{opt}$  as the smoothing constant, we redo the recursive calculation of the smoothed volatility estimate, and the last estimate  $S_{t-1}$  is used as the volatility forecast  $\hat{\sigma}_t$  for  $v_t$ .

The DD-EWMA volatility forecasting model is shown to have a smaller asymptotic variance than conventional estimators and is more appropriate for financial data with larger kurtosis [7]. By using this model to estimate the volatility of the Kalman Filter innovations, the filtering algorithm's stability is improved [2].

## 2.4 Hidden Markov Model and Regime-Aware Thresholds

A Hidden Markov Model (HMM) is a statistical model used to represent sequences of observations, where the underlying state of the system generating the observations is not directly observable. The model assumes that the states generating the observations form a Markov process and that the observations depend on the underlying states.

In this project, the HMM is used to model the market regime, with two possible states: possibly "normal" and "extreme". The model uses probabilities to estimate the likelihood of the underlying state sequence based on the observed sequence. At each time  $t$ , the observed variables  $X_t$  are the innovation  $v_t$  and the returns of each of the two stocks. Therefore,  $X_t = [v_t \ P_{1,t}/P_{1,t-1} - 1 \ P_{2,t}/P_{2,t-1} - 1]$ . The latter two variables, not included in the original paper, are considered because of their intuitive relevance to regime detection. The model assumes that future hidden states depend only on the current hidden state, as determined by the Markov assumption. The state transition probability matrix  $\Pr = p_{ij}$  where  $p_{ij}$  represents the probability of transitioning from state  $i$  to state  $j$ .

To fit the HMM, the transition matrix and emission probabilities must be estimated using the expectation-maximization (EM) algorithm. Define the transition probability matrix of hidden states as:

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

Here,  $i, j \in \{1, \dots, K\}$ . Observable data  $X_t$  is linked to the hidden state  $S_t$  by emission probabilities  $\Pr(X_t | S_t)$ . The joint density of the hidden states and observable data is given as:

$$\Pr(X_{1,\dots,n}, S_{1,\dots,n}) = \Pr(S_1) \prod_{t=2}^n \Pr(S_t | S_{t-1}) \prod_{t=2}^n \Pr(X_t | S_{t-1})$$

The algorithm estimates the probability of being in each state at each time point, as well as the mixture component for each state. The parameters for the emission probabilities,  $\mu$  and  $\Sigma$ , are updated using the following equations:  $\mu_{il} = \frac{\sum_{t=1}^n \gamma_{il,t} X_t}{\sum_{t=1}^n \gamma_{il,t}}$ ;  $\Sigma_{il} = \frac{\sum_{t=1}^n \gamma_{il,t} (X_t - \mu_{il})(X_t - \mu_{il})^T}{\sum_{t=1}^n \gamma_{il,t}}$ , where  $\gamma_{il,t}$  is the joint probability of being in state  $i$  and having observation  $l$  at time  $t$ .

To generate trading signals, a dynamic fitting of the trading threshold ( $p$ ) is used, with an optimal threshold

value  $p_{opt}(S_t)$  determined for each hidden state  $S_t = j, j \in 1, \dots, K, t = 1, \dots, n$ . The upper and lower trading bands are calculated using  $p_{opt}(S_t)\hat{\sigma}_t$ , where  $\hat{\sigma}_t$  is the estimated volatility at time  $t$ . The optimal thresholds are determined from the training dataset by optimizing various statistics such as the Sharpe Ratio, the hit rate, the average win and loss, and the average return, using a brute-force grid search method. Each of the statistics is described in the next subsection.

## 2.5 Trading Strategy and Simulation

The trading strategy involves several steps. First, we use the Engle-Granger two-step approach to identify a cointegrated pair of stocks. After an initialization period, the trading horizon begins. At the end of each day in the trading horizon, we observe the closing prices of the stocks. Using the KF, we obtain the innovation  $v_t$  and the predicted hedge ratio  $\hat{\beta}_1(t+1|t)$ . We estimate the volatility  $\hat{\sigma}_t$  of the innovation using DDIVF, and we use the fitted HMM to detect the current market regime  $S_t$ . Next, we generate a trading signal based on the observed innovation, the estimated volatility, and the current regime. We generate a sell signal when  $v_t$  crosses  $p_{opt}(S_t)\hat{\sigma}_t$  from below and a buy signal when  $v_t$  crosses  $-p_{opt}(S_t)\hat{\sigma}_t$  from above. If  $v_t$  does not cross these thresholds, we generate no signal. After generating the trading signal, we trade a portfolio that exposes us to the spread. Specifically, the portfolio includes a long position in stock 2 and a short position in stock 1 with a position size equal to the current hedge ratio  $\hat{\beta}_1(t|t)$  times our investment in stock 1. It is clear that each signal initiates a one-day position on the portfolio. Finally, we realize the profits and losses based on the daily returns of the stocks and our position on the previous day. All of these actions occur at the end of each trading day during the trading horizon.

To simulate the trading more rigorously, let's denote the following.

- $Q_{i,n}$ : Investment of Stock  $i$  at the start of period  $n$ .
- $R_{i,n}$ : Return of Stock  $i$  over period  $n$ .
- $r_n$ : Fed Fund Rate over period  $n$ .
- $r + \delta r$ : Interest paid for cash on long stock.
- $r - \delta r$ : Interest received for cash on the short stock.
- $\xi$ : Market impact, clearing, and commissions.
- $E_n$ : Equity in the account at the start of period  $n$ .
- $\Lambda_{max}$ : Maximum leverage ratio.

The trading profit and loss ( $\Delta PnL_n = E_{n+1} - E_n$ ) in each trading day are calculated as follows.

$$E_{n+1} - E_n = r_n \Delta t E_n + \sum_{i=1}^N Q_{i,n} R_{i,n} - r_n \Delta t \sum_{i=1}^N Q_{i,n} - \delta r \Delta t \sum_{i=1}^N |Q_{i,n}| - \xi \sum_{i=1}^N |Q_{i,n+1} - Q_{i,n}| \quad (1)$$

Here,  $\sum_{i=1}^N |Q_{i,n}| \leq \Lambda_{max} E_n$ .

The trading rule is stated as follows. I assume without loss of generality that  $E_0 = 100000$ ,  $\xi = 0.0005$ ,  $\delta r = 0$ ,

and  $\Lambda_{max} = 2$ . The trading frequency of this strategy is one trading day  $\Delta t = 1/252$ . Set a constrain that  $\sum_{i=1}^N |Q_{i,n}| = \min(200000, \Lambda_{max} E_n)$  to constrain the exposure. Once the equity account lost all of its value, it would be replenished to the original value, until the cumulative loss exceeded the times the original value.

The following statistics are obtained to optimize the trading thresholds (training set) and evaluate the trading performance (training set and testing set) and are defined as follows.

- Average return:  $\mu = \frac{1}{\Delta t N} \sum_{n=1}^N \frac{E_n - E_{n-1}}{E_{n-1}}$ , where  $N$  is the number of periods over the trading horizon.
- Sharpe Ratio:  $S = \frac{\sum_{n=1}^N (\frac{E_n - E_{n-1}}{E_{n-1}} - r_n)}{\sum_{n=1}^N (\frac{E_n - E_{n-1}}{E_{n-1}} - \mu \Delta t)^2}$ .
- Number of trades: Here, the number of trades equals the number of signals that are not zero over the trading horizon.
- Hit Rate: The number of successful trades divided by the number of successful trades. A successful (failed) trade is defined as a trade that generates positive (negative) profit and loss  $\Delta P n L_t$ .
- Average win: The average of the profit and loss  $\Delta P n L_t$  generated by successful trades.
- Average loss: The average of the negative profit and loss  $-\Delta P n L_t$  generated by failed trades.

Finally, a buy-and-hold strategy for each of the stocks in the pair is simulated according to the same trading rule for comparison.

## 3 Results

### 3.1 Data

This project applies the proposed strategy to the major stocks in the utility sector, namely NEE (NextEra Energy Inc.), D (Dominion Energy Inc.), DUK (Duke Energy Corporation), ED (Consolidated Edison Inc.), AWK (American Water Works Company Inc.), AOS (A.O. Smith Corporation), XEL (Xcel Energy Inc.), SO (Southern), AEP (American Electric Power Company), and SRE (Sempra Energy). The Federal 13-week T-bill Rates are used to calculate the daily risk-free rate.

The dataset is obtained from Yahoo Finance for the period between January 1st, 2017 and March 15th, 2023. The training set consists of data from January 1st, 2017 to December 31st, 2019, while the testing set consists of data from January 1st, 2020 to March 15th, 2023. Figures 1, 2, and 3 show the prices, distribution, and pairwise correlation of the pool of stocks, respectively.

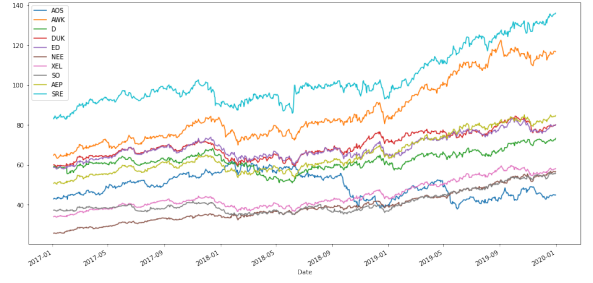


Figure 1: Prices of the Stock Pool: Overview

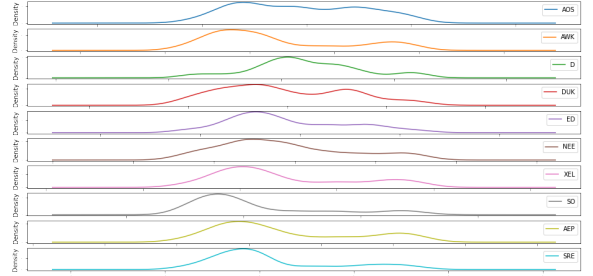


Figure 2: Prices of the Stock Pool: Distribution

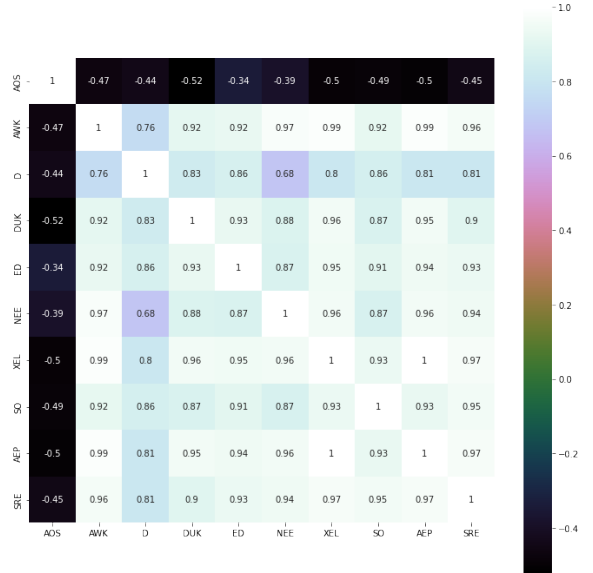


Figure 3: Prices of the Stock Pool: Pair-Wise Correlations

### 3.2 Cointegrated Pairs Selection

The twelve pairs of stocks with p-values obtained from the Engle-Granger two-step approach below 10% are selected for back-testing. These pairs are AOS & DUK (8.77%), AWK & ED (2.71%), AWK & XEL (1.50%), AWK & AEP (0.87%), AWK & SRE (6.18%), D & ED (3.31%), ED & SRE (4.14%), NEE & SRE (8.74%), XEL & AEP (0.01%), XEL & SRE (0.39%), SO & SRE (0.43%), and AEP & SRE (0.76%). The strategy is demonstrated using the AOS DUK pair (figure 4) in accordance with the paper, which was the best-performing pair in the study.



Figure 4: Demnstrative Stock Pair - AOS & DUK: Stock Prices

### 3.3 KF-DDIVF-HMM

The Kalman-filtered beta estimates of the AOS DUK pair are shown in Figure 5. The spread appears to have the momentum to increase. The hedge ratio estimation is insensitive to the assumptions on measurement noise and process noise (Figure 6) even when  $\delta$  and  $\sigma_e^2$  increase by 10,000%. The innovations and their DDIVF forecasted volatility are shown in Figure 7. The innovation marked with HMM detected states are shown in Figure 8, with the innovations more diverged in state zero than state one.

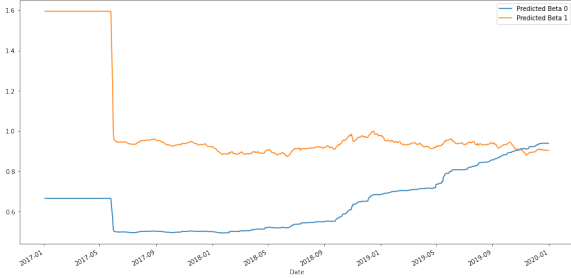


Figure 5: Betas Estimation with KF



Figure 6: Betas Estimation with KF: Sensitivity Test

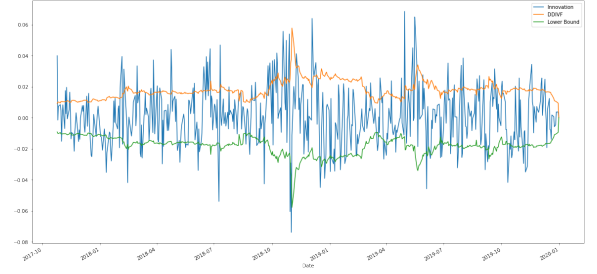


Figure 7: Innovations and Forecasted Volatility with KF-DDIVF

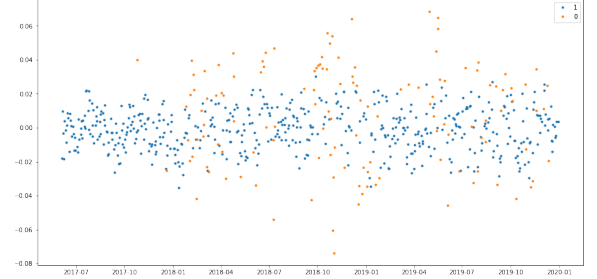


Figure 8: Innovations Marked with States with HMM

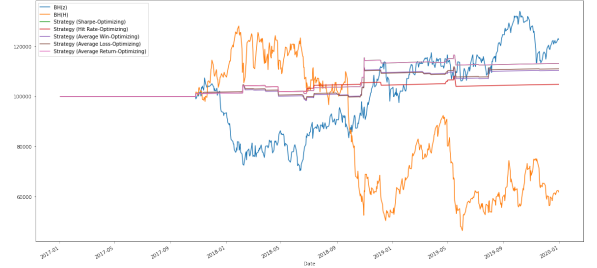


Figure 9: Training Set Performance (AOS & DUK)

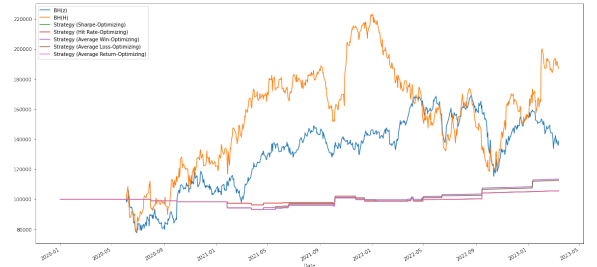


Figure 10: Testing Set Performance (AOS & DUK)

### 3.4 Thresholds Optimization and Training Set Performance

Five sets of optimized thresholds that maximize five trading performance statistics (Sharpe Ratio, average return, hit rate, average win, and average loss) are obtained using grid search. The trading statistics based on the training data are shown in Table 1, and Figure 9 shows the simulated trading results.

### 3.5 Testing Set Performance

The trading statistics based on the testing data are shown in Table 2, and Figure 10 shows the simulated trading results. All strategies outperform the buy-and-hold strategy based on Sharpe Ratio, with the average-win optimizing strategy performing the best.

### 3.6 Performance Robustness Examination

To test the robustness of the strategy performance, I simulate trading based on the other eleven cointegrated stock pairs and obtain the testing set performance statistics. The results of the average-win-optimizing strategy are shown in Table 3. The other four strategies perform just as poorly.

The results show that the strategy is ineffective when applied to other pairs of stocks, consistently underperforming in almost all cases.

Table 1: Training Set Performance (AOS & DUK)

Training Set Performance: AOS & DUK								
	$p_{opt}(S_0)$	$p_{opt}(S_1)$	#Trades	Hit Rate	Avg. W.	Avg. L.	Avg. R.	Sharpe
Sharpe Optimizing	2.2	2.0	18	56%	2167	1222	5.9%	10.25
Hit Rate Optimizing	2.4	2.0	12	67%	1322	1862	2.2%	5.83
Avg. Win Optimizing	1.8	2.4	27	37%	2430	833	4.8%	6.41
Avg. Loss Optimizing	1.8	1.8	30	40%	2161	824	5.1%	6.82
Avg. Ret. Optimizing	2.2	2.0	18	56%	2167	1222	5.9%	10.25
BH (AOS)								-0.49
BH (DUK)								1.54

Table 2: Training Set Performance (AOS & DUK)

Testing Set Performance: AOS & DUK								
	$p_{opt}(S_0)$	$p_{opt}(S_1)$	#Trades	Hit Rate	Avg. W.	Avg. L.	Avg. R.	Sharpe
Sharpe Optimizing	2.2	2.0	17	53%	1787	1365	2.1%	5.47
Hit Rate Optimizing	2.4	2.0	15	47%	1860	984	2.1%	7.67
Avg. Win Optimizing	1.8	2.4	20	65%	1738	1352	4.7%	13.75
Avg. Loss Optimizing	1.8	1.8	23	61%	1710	1262	4.4%	12.12
Avg. Ret. Optimizing	2.2	2.0	17	53%	1787	1365	2.1%	5.47
BH (AOS)								1.68
BH (DUK)								1.55

Table 3: Performance Robustness Examination (Average Win-Optimizing strategy)

Performance Robustness Examination (Average Win-Optimizing strategy)								
	$p_{opt}(S_0)$	$p_{opt}(S_1)$	#Trades	Hit Rate	Avg. W.	Avg. L.	Avg. R.	Sharpe
AWK & ED	2.6	2.6	11	45%	1360	883	0.9%	2.81
AWK & XEL	1.0	2.4	45	49%	854	949	-1.9%	-12.37
AWK & AEP	2.4	2.0	13	38%	933	806	-0.4%	-19.06
AWK & SRE	1.0	2.2	47	57%	831	1395	-2.7%	-7.74
D & ED	2.0	2.6	16	31%	792	595	-0.8%	29.80
ED & SRE	2.2	2.8	11	36%	898	1068	-0.11%	-27.92
NEE & SRE	2.2	1.2	42	45%	1265	966	0.0%	-2.10
XEL & AEP	2.0	1.4	38	61%	579	755	0.1%	-6.51
XEL & SRE	2.4	2.8	13	54%	1050	1516	-0.3%	-8.20
SO & SRE	2.4	2.4	10	30%	903	881	-0.9%	-29.78
AEP & SRE	2.8	2.6	7	29%	1101	775	-0.1%	-25.70

## 4 Discussions

This strategy is tested to be generally ineffective and unprofitable when applied to stocks in the utility sector from 2021 to 2023. One possible explanation for this finding is the momentum of the spread divergence. The profitability of the strategy

relies on the assumption that the spread will return to equilibrium within one day after it diverges significantly. However, in reality, the divergence of the spread may have a certain inertia even when it is significant enough. Therefore, one potential improvement to the strategy could be to model the inertia of the spread divergence or assume that the spread exhibits either mean-reverting or momentum behavior. Further research is needed to explore these possibilities and improve the effectiveness of the strategy when applied to other cointegrated stock pairs. Another potential issue is the overfitting of the training set. The profitability of the strategy was found to be highly sensitive to the thresholds used to generate signals.

## References

- [1] Johnson-Skinner, E., Liang, Y., Yu, N., Morariu, A. (2021, July). A Novel Algorithmic Trading Strategy using Hidden Markov Model for Kalman Filtering Innovations. In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 1766-1771). IEEE.
- [2] Liang, Y., Thavaneswaran, A., Hoque, M. E. (2020, December). A Novel Algorithmic Trading Strategy Using Data-Driven Innovation Volatility. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1107-1114). IEEE.
- [3] Engle, R. F., Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.
- [4] Elliott, R. J., Van Der Hoek\*, J., Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance*, 5 (3), 271-276.
- [5] De Moura, C. E., Pizzinga, A., Zubelli, J. (2016). A pairs trading strategy based on linear state space models and the Kalman filter. *Quantitative Finance*, 16 (10), 1559-1573.
- [6] Becker, A. (n.d.). Online Kalman filter tutorial. Retrieved 2023, from <https://www.kalmanfilter.net/default.aspx>.
- [7] Thavaneswaran, A., Paseka, A., Frank, J. (2020). Generalized value at risk forecasting. *Communications in Statistics-Theory and Methods*, 49 (20), 4988-4995.