Cointegration Approach for the Pair Trading based on the Kalman Filter

Jia Yu

Graduate School of Information, Production and Systems, Waseda University, Fukuoka, Japan

joanna-yu@suou.waseda.jp

Abstract. The conventional pairs trading strategy of cointegration approach employs a fixed cointegration coefficient, limiting the length of the applicable trading period and negatively influencing profitability. Given these two problems, this paper proposes a pairing trading system based on the co-integration method of the Kalman filter, which can provide investors with profitable trading pairing and market timing. First of all, study the co-integration relationship between every two stocks and co-integrate the stock as a trading pair. Then, based on the analysis of the trading value of the trading stock price by the state-space model, a Kalman filter system to update the co-integer coefficient value is designed. Finally, a trading strategy is developed for the proposed method. In addition, simulations were conducted to assess the performance of the proposed trading system. According to the conventional method and the recommended trading method, twenty pairs of cointegrated stocks were selected to build trading pairs respectively. The simulation results prove that the proposed approach has higher profitability and a more extended trading period than the conventional approach.

Keywords: Pairs Trading Strategy; Cointegration Approach; Kalman Filter; Time Series Analysis; State-space Model

1 Introduction

Pairs trading is a statistical arbitrage first introduced by Nunzio Tartaglia, designed to identify, and exploit market inefficiencies observed by two long-term related assets, primarily through statistical methods [1]. The co-integration method of pairing transactions was proposed by Vidyamurthy, who proposed the use of a mathematical relationship between two stocks, i.e., a co-integration relationship, to build a trading pair [2]. Afterwards, this cointegration framework was employed by Hossein et al. to construct a complete trading strategy [3]. In this strategy, Hossein et al. defined a 6-month trading period and a 12-month parameters' formation period, which means the strategy's parameters in the trading period were estimated in the formation period. Therefore, the cointegration coefficients of trading pairs were fixed values calculated by data from the formation. However, the value of the cointegration coefficients changes along

V. Escudero et al. (Eds.): BADS 2022, AHCS 11, pp. 633–642, 2023.
https://doi.org/10.2991/978-94-6463-102-9_66

with the changes in stock prices of the two selected stocks. Consequently, the accuracy of the cointegration coefficients employed by Hossein et al. is low, which will negatively influence the profitability of the trading strategy.

To solve this problem, the Kalman Filter, proposed by R.E. Kalman, is employed in this paper to find a better way to calculate and update the value of the cointegration coefficient [4]. Also, this paper aims to propose a better pairs trading strategy with higher investment returns for investors.

2 Pairs Trading based on Cointegration Approach

A pairing strategy is a market-neutral trading strategy that matches long and short positions between two highly relevant stocks. These two highly related stocks are defined as trading pairs. When a significant deviation in the price between the two assets is detected, a trading position is taken: sell the higher-priced asset, buy the lower-priced asset, and expect the mispricing to correct the long-term equilibrium value [1]. Cointegration is one of the methods of selecting trading pairs and constructing pair trading strategies [3].

For a non-stationary time series X_t which is called $I_{(1)}$, if its first differential forms a stationary process, i.e. $I_{(0)}$. Consider $X_{1,t}$ and $X_{2,t}$ to be two $I_{(1)}$ time series. A linear combination of the two-time series is stationary, $X_{1,t}$ and $X_{2,t}$ are said to be cointegrated. Therefore, $X_{1,t}$ and $X_{2,t}$ are cointegrated if there exists a non-zero real number β such that

$$X_{2,t} - \beta X_{1,t} = u_t \tag{1}$$

β is the cointegration coefficient, and u_t is a stationary series known as cointegration errors. According to Hossein et al., the spread series of the trading pairs is defined as the scaled difference in the prices of two stocks:

$$Spread_t = X_{2,t} - \beta X_{1,t} \tag{2}$$

Besides, from (1), the spread is stationary by definition and shows the mean-reverting property based on what Hossein et al. constructed their trading strategy. In this trading strategy, a 6-month trading period is used to execute the strategy using the cointegration coefficient β estimated in the previous 12 months, defined as the formation period. Hossein et al. calculated the cointegration coefficient β by (3), $X_{1,t}$ and $X_{2,t}$ are the stock prices of stock 1 and stock 2 from the formation period:

$$\beta = \frac{\sum X_{1,t}X_{2,t} - n\overline{X_{1,t}X_{2,t}}}{\sum X_{1,t}^2 - n\overline{X_{1,t}}} \tag{3}$$

This means that the cointegration coefficient's fixed value is applied to a trading cycle and updated every six months to ensure its accuracy. In practice, however, the value of the cointegration coefficient is a dynamic value that changes over the life of the entire trading cycle as the two co-integrated stock prices change. To clarify this phenomenon, Fig.1 shows the value of the cointegration coefficient β of a trading pair in reality β and the value calculated by the method of Hossein et al., in a 6-month formation period and an 18-month trading period. Thus, Fig. 1 shows the value of β calculated by daily stock prices in a 1278-day period. As shown by Fig.1, the value of β used by Hossein et al. is at low accuracy, which will negatively influence the profitability of the trading strategy. In order to solve this problem, the spatial state model of financial time series is used to build the model. The Kalman Filter is employed to update the value of the cointegration coefficient in the real-time, therefore, to propose a pairs trading strategy with higher profitability.

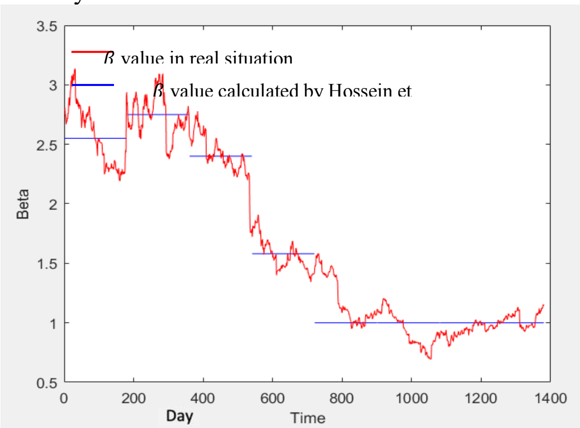

Fig. 1. Value of cointegration coefficient β

3 The Proposed Trading Strategy

3.1 Framework

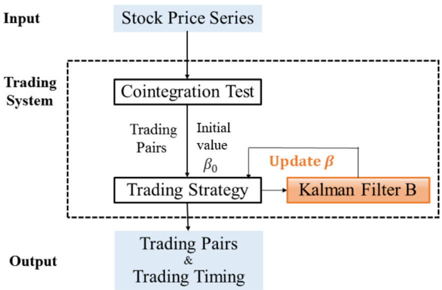

Fig. 2. Proposed Trading System

The proposed trading system is shown in Fig.2. The input of the trading system is the stock-price series. The Johansen Test [3] is applied as the method of the cointegration test. Next, based on the cointegration test results, every two cointegrated stocks will be grouped as a trading pair that will also be the output of the system. The initial value of the cointegration coefficient β_0 is calculated for each trading pair. The Kalman Filter will update the value of the cointegration coefficient in real-time to improve its accuracy. The most recently updated value of the cointegration coefficient will be used to calculate the $Spread_t$, which is used as the trading signal in the trading strategy. When the $Spread_t$ meets defined condition, a trade will be made, and the corresponding time t of $Spread_t$ is outputted as the trading timing. In this way, the proposed system can provide selected trading pairs and optimal trading timing for investors.

3.2 Kalman Filter System

R.E.Kalman proposed the Kalman Filter system based on the state-space model in 1960 [6]. Durbin, J. and Koopman, S. J. stated that stock prices could also be deemed financial time series and presented by the state-space model [7]. For the two cointegrated stocks, their daily close price is used to construct the two stock price series $X_{1,t}$ and $X_{2,t}$, which satisfy the following equation:

$$Spread_t = X_{2,t} - \beta_t X_{1,t}. \tag{2}$$

Therefore,

$$X_{2,t} = \beta_t X_{1,t} + Spread_t \text{ (Measurement Equation)} \tag{4}$$

As the property of the cointegration relationship, the series of $Spread_t$ means reverting to the value, zero. Assume that the series $Spread_t \sim N(0, Q)$ and Q are the covariance for the series [8]. Therefore, $Spread_t$ can be considered a Gaussian noise series, and (2) can be considered a measurement equation in Kalman Filter systems. The goal of the Kalman filter system is to update the value of the co-integration coefficient β_t, so assume that the co-coefficient at the time *t* is equal to the co-coefficient at the *t*-1 moment plus noise:

$$\beta_t = \beta_{t-1} + v_t \text{ (State Equation).} \tag{5}$$

v_t is also a Gaussian noise with covariance R. Thus, $v_t \sim N(0, R)$.Based on the state equation and observation equation described above, the system of the discrete Kalman Filter can be built. The following equations can describe the estimation and update processes for this system:

$$\hat{\beta}_{t|t-1} = \hat{\beta}_{t-1|t-1} \tag{6}$$

$$P_{t|t-1} = P_{t-1|t-1} + R \tag{7}$$

$$P_{t|t-1} = cov(\beta_t - \hat{\beta}_{t|t-1}) \tag{8}$$

$$\hat{X}_{2,t} = X_{1,t}\hat{\beta}_{t|t-1} \tag{9}$$

$$Q = cov(Spread_t) \tag{10}$$

$$G_t = X_{1,t}^T P_{t|t-1} X_{1,t} + Q \tag{11}$$

$$Spread_t = X_{2,t} - X_{1,t}\hat{\beta}_{t|t-1} \tag{12}$$

$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + K_t Spread_t \tag{13}$$

$$P_{t|t} = P_{t|t-1} - K_t X_{1,t} P_{t|t-1} \tag{14}$$

$$K_t = P_{t|t-1} X_{1,t} / G_t \tag{15}$$

Where
$\hat{\beta}_{t|t-1}$ *is the estimate value of* β *at time t based on the observation at time* $t-1$.
$\hat{\beta}_{t|t}$ *is the estimate value of* β *at time t based on the observation at time t* .
$\hat{X}_{2,t}$ *is the estimate value of* $X_{2,t}$ *at time t based on the observation at time* $t-1$.
P_t *and* G_t *are covariance matrices, and assume* $P_{0|0} = 0$.
K_t *is the Kalman gain.*

The value of β_0 is calculated using the historical data of the stock price series. The covariance Q is defined as the variance of the series of spreads calculated by actual stock prices and cointegration coefficient β at each sampling point. Moreover, the covariance R is defined as the covariance of the volatility of the series of β calculated by the equilibrium prices of two cointegrated stocks. In order to obtain the equilibrium price of the stock, another discrete Kalman Filter system is designed. The stock price series can be seen as a signal with noise and the short-term deviation can be deemed as the noise. Meanwhile, the long-term equilibrium price can be considered as the price that the noise has been removed from the actual stock price. Furthermore, a Kalman Filter is designed to remove the noise of the actual stock price series.

The second Kalman Filter is built based on the theory provided by Eugene [5], which is that the stock price series is modelled as a random walk. Therefore, a stock price series can be presented by (16), which is employed as the state equation of the Kalman Filter system.

$$x_t = x_{t-1} + w_t \text{(State Equation)} \tag{16}$$

$$w_t \sim N(0, V_w)$$

Where
x_t *is the actual stock price at time* t.

w_t *is a Guassian noise series.*

V_w *is the covariance of the* w_t.

$V_w = variance\ of\ the\ period\ returns$

Assume

$$y_t = x_t + v_t (\text{Measurement Equation}) \tag{17}$$

$$v_t \sim N(0, V_v)$$

Where
y_t *is the equilibirum stock price at time t.*

v_t *is a Guassian noise series.*

V_v *is the covariance of the* $v_{t,}$

$V_v = variance\ of\ the\ period\ volatilities$

Equation (17) is used as the measurement equation of the Kalman Filter. The historical price of stocks calculates the V_w and V_v. Moreover, the outputs of this Kalman Filter are the equilibrium prices of stocks.

If the equilibrium prices of $X_{1,t}$ *and* $X_{2,t}$ are $Y_{1,t}$ *and* $Y_{2,t}$, respectively, the corresponding cointegration coefficient $\beta_{(e)t} = regress(Y_{1,t}, Y_{2,t})$. Then, the volatility series is $\Delta\beta_{(e)t} = \beta_{(e)t} - \beta_{e(t-1)}$. Therefore,

$$\mathrm{R} = cov(\Delta\beta_{(e)t}) \tag{18}$$

3.3 Trading Strategy

The trading strategy consists of two parts: the selection of trading pairs and selecting the timing of trading. As for the selection of trading pairs, only cointegrated stocks are selected to construct the trading pairs. In this part, the Johansen test is employed to identify whether there is a cointegration relationship between two stocks.

To use the spread as the trading signal, the spread is normalized by (19):

$$Normalized\ Spread_t = \frac{Spread_t - Mean\ of\ Spread_t}{Standard\ Devation\ of\ Spread_t}$$

$$= \frac{Spread_t}{Standard\ Devation\ of\ Spread_t} \tag{19}$$

-1 and +1 are selected as the threshold for the normalized spread, because when the normalized spread moves out of the range [-1, 1], it means that two stocks move away for their long-term equilibrium relationship of the normal situation. In other words, in these situations, the short-term deviation can be used to make profits. Therefore, if the shareholder does not hold any stock, when the normalized spread drops below -1, buy one dollar worth of stock 2 and sell β dollar worth of stock 1 at the same time; otherwise, hold the stocks on the hand. If the shareholder does not hold any stock, when the normalized spread moves above +1, sell *1/*β dollar worth of stock 2 and buy one dollar worth of stock 1 simultaneously; otherwise, hold the stocks at hand. When the normalized spread returns to zero, close both positions.

4 Simulation

4.1 Data

The dataset for simulation consists of daily data in the New York Stock Exchange stocks from 1st September 2011 to 28th February 2017. The close prices for the stocks are used, and the data-set sample period is 1381 days (66 months). The formation period is from 1st September 2011 to 29th February 2012. The data from the formation period is used to calculate the covariance Q. The trading period is from 1st March 2012 to 28th August 2017, and 20 trading pairs are selected from the total stocks by the cointegration test, as shown by Tab.1.

Table 1. Selected Trading Pairs

No.	**Stock Name of Trading Pairs**	
	Stock 1	***Stock 2***
1	Advance Auto Parts, Inc. (AAP)	International Business Machines Corporation (IBM)
2	iShares MSCI Australia ETF (EWA)	iShares MSCI Canada ETF (EWC)
3	NIKE, Inc. (NKE)	The AES Corporation (AES)
4	Merck & Co., Inc. (MRK)	Verizon Communications Inc. (VZ)
5	The Goldman Sachs Group, Inc. (GS)	Chevron Corporation (CVX)
6	Merck & Co., Inc. (MRK)	E. I. du Pont de Nemours and Company (DD)
7	The AES Corporation (AES)	Exxon Mobil Corporation (XOM)
8	Chevron Corporation (CVX)	Affiliated Managers Group, Inc. (AMG)
9	Verizon Communications Inc. (VZ)	The Coca-Cola Company (KO)
10	Verizon Communications Inc. (VZ)	Johnson & Johnson (JNJ)
11	Verizon Communications Inc. (VZ)	The Procter & Gamble Company (PG)
12	The Coca-Cola Company (KO)	Wal-Mart Stores, Inc. (WMT)
13	Wal-Mart Stores, Inc. (WMT)	Visa Inc. (V)
14	General Electric Company (GE)	International Business Machines Corporation (IBM)
15	Wal-Mart Stores, Inc. (WMT)	Exxon Mobil Corporation (XOM)
16	The Travelers Companies, Inc. (TRV)	Exxon Mobil Corporation (XOM)
17	The Procter & Gamble Company (PG)	E. I. du Pont de Nemours and Company (DD)
18	NIKE, Inc. (NKE)	Verizon Communications Inc. (VZ)
19	Cisco Systems, Inc. (CSCO)	Exxon Mobil Corporation (XOM)
20	Accenture plc (CAN)	Verizon Communications Inc. (VZ)

To avoid relatively high trading costs and complications, these selected stocks are restricted to be liquid stocks. This is done by removing the bottom decile stocks, in terms of market cap, in the sample period. And for the same reason, stocks with prices less than $1 in the sample period are also not considered. Besides, different brokers in the

New York Stock Exchange charge different transaction costs, which means the same transaction may be charged different transaction costs, so the transaction costs are not considered in the simulation process.

4.2 Simulation Results and Discussion

The profitability is used to evaluate the proposed strategy's performance compared to Hossein et al.'s trading strategy.-Moreover, the results of accumulated profit (total profit made from investment in the trading period) are shown in Tab.2. On the one hand, the trading period is enlarged by the proposed approach from 6 months to 5 years. On the other hand, the proposed approach is a better way to decide the cointegration coefficient β than the conventional approach. Meanwhile, the systemic risk can also be better hedged by using a better cointegration coefficient β. As a result, the proposed approach has higher profitability than the conventional approach in a long trading period, at least one year, when the cointegration coefficient β changes dramatically.

Table 2. Accumulated Profit Rate of Investment

No.	Accumulated Profit Rate of Investment		
	Strategy of Hossein et al.	*Proposed Strategy*	*Improvement Rate*
1	107.78%	117.61%	9.12%
2	110.81%	115.37%	4.12%
3	124.28%	125.40%	0.92%
4	108.70%	112.40%	3.43%
5	92.86%	99.73%	7.40%
6	112.47%	114.44%	1.80%
7	115.43%	123.81%	7.31%
8	123.55%	134.38%	8.73%
9	113.59%	116.09%	2.25%
10	101.11%	110.35%	9.16%
11	109.43%	114.11%	4.29%
12	105.47%	105.98%	0.52%
13	105.53%	108.96%	3.22%
14	95.04%	103.49%	8.92%
15	123.82%	129.36%	4.52%
16	111.74%	118.61%	6.12%
17	87.12%	88.04%	1.05%
18	113.26%	118.60%	4.71%
19	100.19%	103.95%	3.76%
20	119.39%	125.53%	5.11%

5 Conclusion

In this paper, a new pairs trading system, which can select the stocks that should be traded and provide optimal timing of trading, is proposed based on the Kalman Filter theory. For the proposed trading system, the inputs are daily close prices of stocks. These inputs are used to identify whether the cointegration relationship exists between the stocks by Johansen Test. Next, each of the two co-integrated stocks is grouped into one trading pair. Then, based on the linear state-space model of the stock price series, the Kalman filter system is constructed to update the dynamic parameter values in real-time: the co-integration coefficient β. After that, the spread is calculated by applying the recently updated co-integration factor β value, which is used as a trading signal in the recommended trading system. When the spread exceeds the range defined by the predetermined threshold, the corresponding trading decisions, including the number of shares sold/bought and the time to complete the trade, are given as the output of the recommended trading system. Furthermore, according to the simulation results, the proposed trading system provides a better way to determine the real-time value of the co-integration coefficient. In the longer trading cycle, the traditional trading strategy has higher profitability in general.

However, there are still some points that can be furtherly improved. For example, trading signals can be further improved by using dynamic thresholds instead of the fixed thresholds used in this work. That could further improve the profitability of the trading strategy. In addition, transaction costs are not included in this work, so for future studies, researchers can take transaction costs into account when structuring trading strategies to bring them closer to the actual situation.

References

1. Gatev, E., Goetzmann, W. N., and Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies, 19(3):797–827
2. Vidyamurthy, G., Pairs Trading: Quantitative Methods and Analysis, 2004 (Wiley: Hoboken, NJ)
3. Hossein Rad, RKY Low, R Faff, The profitability of pairs trading strategies: distance, cointegration and copula methods, Quantitative Finance, 2016, 1541–1558
4. Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. SIGGRAPH, ACM, 2001
5. Eugene F. Fama, Random Walks in Stock Market Prices, Financial Analysts Journal, 1995, 75-80
6. R.E.Kalman, A new approach to linear filtering and prediction problems, ASME Journal of Basic Engineering, 1960, 82:35–45
7. Durbin, J., & Koopman, S. J., Time series analysis by state space methods, Oxford university press, 2012
8. D. Simon, Optimal State Estimation: Kalman, H Infinty, and Nonlinear Approaches. Wiley-Interscience, 2006