

Statistical Arbitrage in S&P500

Stefanos Drakos

International Centre for Computational Engineering, Rhodes, Greece

Email: stefanos.drakos@gmail.com

Received 8 January 2016; accepted 24 February 2016; published 29 February 2016

Copyright © 2016 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A methodology to create statistical arbitrage in stock Index S&P500 is presented. A synthetic asset based on the cointegration relationship of the stocks with Index was constructed. In order to capture the dynamic of the market time adaptive algorithms have been developed and discussed. The pair trading strategy was applied in different periods between S&P500 and synthetic asset and the results were evaluated. Different metrics have shown that the Multivariate Kalman Algorithm creates statistical arbitrage in index with much lower Maximum Drawdown and higher profit. The algorithm is neutral as the beta is close to zero and the Sharp Ratio remains high in all cases.

Keywords

Statistical Arbitrage, Mean Reverting, Pair Trading, Kalman Filter, Trading Algorithms

1. Introduction

Financial markets are based on the general trading rule: buy with low price and sell with high price. The aim is the development of strategies with low risk and succeeds this general rule. Pure Arbitrage is a category of strategies with zero risk. As an example we can refer the case of buying and selling a stock at the same time with a different value in two different exchanges. The profit results from the difference in prices, breaking the law of one price.

Another category is the statistical arbitrage which is not risk free at all. Strategies of this type are aimed at the expected gain which is greater than the risk. The profit results from the mispricing of the stocks. To achieve this, one needs to assess whether the price of a stock is overvalued or undervalued relative to the actual value which is really hard to determine. The fundamentals of the stock, the demand of each period and the general economic environment are some of the factors that make the fair value evaluation difficult. Part of this category is the Pair Trading and is presented in several references as in [1]-[10]. The Pairs trading is a strategy which is based on the relative pricing of stocks without actually interested in the true value of them. The relative pricing is based on

the idea that two assets with the same features can be priced about the same price as the basis of the law of one price. When at a given time the stock prices are different then one is overvalued and the other undervalued relative to the actual price. The classic trading Pairs strategy derives from the characteristics of this incorrect pricing (mispricing) between the two stocks.

In the current work instead of the two stocks the mispricing between the index S&P500 and a sub-set of stocks belonging to this is considered. This forms a portfolio consisting of one unit of the Index in long (or short) and the corresponding number γ (Hedge ratio) of subset's stocks in the opposite position. Three methods adopted to determine the number γ (Hedge ratio): a) the ordinary, b) the rolling ordinary least squares (OLS) regression and c) the Kalman filter process as it is presented below. According to the strategy, the spread of the index S&P500 and the stocks prices' combination are computed and when this deviates from its historical average value then the investor bets on the return to the historical with selling and buying respectively the stocks of the portfolio. In practice to construct the synthetic asset is necessary to investigate the appropriate stock exhibiting long relationship between them with the index. The technique used for this purpose is the Cointegration as this presented in [1] according to which two time series $X_t \sim I(1)$ and $Y_t \sim I(1)$ are cointegrated if $aX_t + bY_t \sim I(0)$ for $a, b \neq 0$ and notation $I(d)$ means integrated order d .

In [1] concluded that pairs who cointegrate in sample period behave better in the out-of-sample period than those not cointegrate in the sample period. Based on the previous, the spread of the linear combination of the cointegrated stocks and the S&P500 index have to be stationary process. To analyze this relationship the augmented Dickey-Fuller test (ADF) was used as presented in [2]. According to the strategy the stocks are not restricted to cointegrate in the out-of-sample period but it is an indication that they will present mean-reverting behavior.

Additionally the log of prices used is also instead the prices as in [3]. The main reason is that log-returns are time additive. So, in order to calculate the return over n periods using real returns we need to calculate the product of n numbers: $(1+r_1)(1+r_2)\dots(1+r_n)$.

If r_1 defined as:

$$r_1 = \frac{P_1 - P_0}{P_0} \Rightarrow (1+r_1) = \frac{P_1}{P_0} \quad (1)$$

And

$$\log(1+r_1) = \log\left(\frac{P_1}{P_0}\right) \quad (2)$$

$$\begin{aligned} & \log(1+r_1) + \log(1+r_2) + \dots + \log(1+r_n) \\ &= \log\left(\frac{P_1}{P_0}\right) + \log\left(\frac{P_2}{P_1}\right) + \dots + \log\left(\frac{P_n}{P_{n-1}}\right) \\ &= \log(P_1) - \log(P_0) + \log(P_2) - \log(P_1) + \dots + \log(P_n) - \log(P_{n-1}) \\ &= \log(P_n) - \log(P_0) \end{aligned}$$

Consequently the profit of spread over a period is equal to:

$$y_{t+i} - y_t = \log\left(\frac{P_{t+i}^A}{P_t^A}\right) - \gamma \log\left(\frac{P_{t+i}^B}{P_t^B}\right) \quad (3)$$

The classic application of the trading pair strategy has published in many works in the past and has been applied in practice. In [4] the author went a further step and presented an algorithm for the pair trading of a stock basket cointegrated with the S&P500 using constant cointegrating coefficient with a single value for all stocks belonging to the basket. To capture the real behavior of the markets this work presents new Time Adaptive algorithms using rolling ordinary least squares (OLS) regression and Multivariate Kalman Filter process where the time dependent hedge ratio is computed separately for each of the stocks forming the synthetic asset creating thereby statistical arbitrage conditions in index S&P500 and increasing the strategy performance.

2. Period of Methodology

The proposed methodology and the trading algorithm designed based on that divided in two different spaces. The first refers to the in sample period which is used to make all the appropriate test and construct the synthetic asset and the other to the out of sample period (Trading Period) where the synthetic asset trading based on the specific rules (**Figure 1**).

2.1. In Sample Period

The data of the in sample period used for the synthetic asset construction. Working on a daily data domain a year of closing prices was chosen as the in sample period for determine the set of cointegrated stocks with S&P500 and create the synthetic asset.

Synthetic Asset Construction.

The paper presents different algorithms in order to create statistical arbitrage in S&P500. The S&P500, based on the market capitalizations of 500 large companies equity indices, and many consider it one of the best representations of the U.S. stock market. The initial choice of stocks for cointegration test with S&P500 is made by the S&P100 which is a sub-set of the S&P500, and measures the performance of large cap companies in the United States. Constituents of the S&P100 are selected for sector balance and represent about 57% of the market capitalization of the S&P500 and almost 45% of the market capitalization of the U.S. equity markets. The stocks in the S&P100 tend to be the largest and most established companies in the S&P500 (*Wikipedia*).

Using the data of a selected In Sample Period for each stock $S_i \in \text{S\&P 100}$ the second step of Engle and Granger' approach adopted. Using the logarithmic price of stocks and S&P500 the OLS regression shown below is performed:

$$\log(S_t^{SPY}) = \gamma_i \log(S_t^i) + \varepsilon_t \quad (4)$$

The Augmented Dickey-Fuller unit root test applied for the stationarity of the OLS residuals. According to the successful stationary result a subset of the S&P100 is created and the components of this set are the candidate for the synthetic asset construction. Still working in the sample period a new OLS regression was carried out:

$$\log(S_t^{SPY}) = \gamma \sum_i^N \log(S_t^i) + \varepsilon_t \quad (5)$$

Or

$$\log(S_t^{SPY}) = \gamma \log\left(\prod_i^N S_t^i\right) + \varepsilon_t \quad (6)$$

Using again the Augmented Dickey-Fuller the stationarity of the new OLS residuals is examined.

If the stationarity exist then this an evidence of mean reverting long term behavior of the spread $\varepsilon_t = \log(S_t^{SPY}) - \gamma \sum_i^N \log(S_t^i)$ where $S_t^i \in S$ and S is the set of stock where individually and as logarithmic sum cointegrated with S&P500. The dimension of S is $\dim(S)$.

2.2. Out of Sample Period-Trading Period

According to the previous period different algorithms for synthetic asset trading are developed. The first was

In Sample Period	Out Of Sample Period
Cointegration Test Synthetic Asset Construction	Trading

Figure 1. Period of study.

designed assumed that the cointegration coefficient is constant during the trading period.

In that case and using the Equation (3) the profit of the strategy during the period $t \div t + h$ arises from the following equation:

$$\varepsilon_{t+h} - \varepsilon_t = \log\left(\frac{S_{t+h}^{SPY}}{S_t^{SPY}}\right) - \gamma \sum_i^N \log\left(\frac{S_{t+h}^i}{S_t^i}\right) \quad (7)$$

3. Time Adaptive Coefficient γ

In reality the system of trading is dynamic and updated as new information get to in and the cointegration coefficient (or the hedge ratio) cannot stay constant during the trading period. For that reason time adaptive algorithms are developed to capture the real conditions of the markets.

3.1. Rolling Ordinary Least Squares (OLS) Regression

The first one considering a rolling ordinary least squares (OLS) regression. The frequency of regression calculations raised by an optimization procedure and the cointegration coefficient calculated at each step by the regression of $\log(S_t^{SPY})$ against the $\log\left(\prod_i^N S_t^i\right)$.

3.2. Kalman Filter Process

The Kalman filter process can be described by three different steps: the prediction the observation and the correction. A new approach were developed using a Multivariate Kalman filter process. Based on that the hedge ratio calculated separately for each stock owned in the synthetic asset and the computed vector of the calculated parameters at each time step has dimensions $(N + 1) \times 1$ where $N = \dim(\mathbf{S})$ while the dimensions of the covariance matrix is $(N + 1) \times (N + 1)$. The aim of these algorithms is to calculate at each time step the updated hedge ratio of the synthetic asset. Assuming that the hedge ratio and the premium follow a random walk we have:

$$y_t = y_{t-1} + w_t \quad (8)$$

where:

y_t : is the current state of the of the parameters.

y_{t-1} : is the previous state of the of the parameters.

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

where for the multivariate Kalman filter process

$$y_t = [\gamma_t^1 \quad \gamma_t^2 \quad \cdots \quad \gamma_t^n \quad \mu_t]^T \quad (9)$$

With:

$$y_0 = [h \quad h \quad \cdots \quad h \quad \mu_0]^T \quad (10)$$

h : is the cointegration coefficient from in sample period and μ_0 coming from the same period.

The vector of logarithmic price of stocks:

$$x_t = [\log(S_t^1) \quad \log(S_t^2) \quad \cdots \quad \log(S_t^n) \quad 1] \quad (11)$$

And $S_t^i \in \mathbf{S}$

The process following the steps as below:

Prediction state where the next system state is predicted based on the knowledge of the previous state

$$\hat{y}_{t|t-1} = y_{t|t-1} + w_t \quad (12)$$

The covariance of prediction state is given by:

$$\hat{P}_{t|t-1} = \hat{P}_{t-1|t-1} + V_w \quad (13)$$

The next step concerns the measurement prediction. Given the price of the synthetic asset and the predicted

hedge ratio the measurement prediction are given as:

$$\hat{z}_t = x_t \cdot \hat{y}_{t|t-1} \quad (14)$$

The residual of measurement and real value at each step calculated as:

$$\varepsilon_t = \log(S_t^{SPY}) - \hat{z}_t \quad (15)$$

The variance of measurement error is equal to:

$$S_t = x_t \cdot \hat{P}_{t|t-1} \cdot x_t^T + V_e \quad (16)$$

The Kalman Gain is the filter, which tells how much the predictions should be corrected on time step is given as:

$$K_t = \frac{\hat{P}_{t|t-1} \cdot x_t^T}{S_t} \quad (17)$$

The last step of process is the update step where:

The updated state is estimated as following:

$$\hat{y}_{t|t} = \hat{y}_{t|t-1} + K_t \cdot \varepsilon_t \quad (18)$$

And the Updated state covariance is equal to

$$\hat{P}_{t|t} = \hat{P}_{t|t-1} - K_t \cdot S_t \cdot K_t^T \quad (19)$$

All the process repeated at every time step of out of sample period. The estimation of V_w and V_e has been discussed in [11] [12].

3.3. Profit of Strategies

In the case of Time Adaptive coefficient γ of the linear regression case the profit of the pair trading strategy raised by the following equations:

$$\varepsilon_{t+h} - \varepsilon_t = \log\left(\frac{S_{t+h}^{SPY}}{S_t^{SPY}}\right) - \log\left(\frac{\gamma_{t+h} \sqrt{\prod_i^N S_{t+h}^i}}{\gamma_t \sqrt{\prod_i^N S_t^i}}\right) \quad (20)$$

In the case of Multivariate Kalman Filter where the hedge ratio is different for each stock and for each time step in the synthetic asset it can be shown that:

$$\begin{aligned} \varepsilon_1 &= \log(S_1^{SPY}) - (\gamma_1^1 \log(S_1^1) + \gamma_1^2 \log(S_1^2) + \dots + \gamma_1^n \log(S_1^n)) \\ &= \log(S_1^{SPY}) - \left(\log\left(\prod_i^N \sqrt[\gamma_i^1]{S_1^i}\right) \right) \end{aligned}$$

Similarly

$$\begin{aligned} \varepsilon_2 &= \log(S_2^{SPY}) - \left(\log\left(\prod_i^N \sqrt[\gamma_i^2]{S_2^i}\right) \right) \\ &\vdots \\ \varepsilon_{t+h} &= \log(S_{t+h}^{SPY}) - \left(\log\left(\prod_i^N \sqrt[\gamma_i^{t+h}]{S_{t+h}^i}\right) \right) \end{aligned}$$

And finally the profit during a period $t/t+h$ is equal to:

$$\varepsilon_{t+h} - \varepsilon_t = \log\left(\frac{S_{t+h}^{SPY}}{S_t^{SPY}}\right) - \log\left(\frac{\prod_i^N \sqrt[\gamma_i^{t+h}]{S_{t+h}^i}}{\prod_i^N \sqrt[\gamma_i^t]{S_t^i}}\right) \quad (21)$$

4. The Pair Trading Strategy

The algorithm of the pair trading strategy is based on the distance of the spread from its historical mean value and its mean-reverting behavior. To measure this distance a normalized variable called z-score introduced as:

$$z_t = \frac{\varepsilon_t - \mathbb{E}[\varepsilon_t]}{\sigma(\varepsilon_t)} \quad (22)$$

where:

$\mathbb{E}[\varepsilon_t]$: is the mean value of the spread over a lookback period.

$\sigma(\varepsilon_t)$: is the standard deviation of the spread over the same period.

The trading it takes place when this variable exceeds some limits based on the spread mean reverting behavior.

Thus:

Open-long position if $z_t < z_{\text{low}}$.

Open-short position if $z_t > z_{\text{high}}$.

Exit-long position if $z_t > z_{\text{exit-low}}$.

Exit-short position if $z_t < z_{\text{exit-high}}$.

When a long (short) position is opened we buy (sell) one unit of S&P500 and sell (buy) the following amount of stocks from synthetic asset as: $\gamma \sum_i^N \log(S_t^i)$ in the case of constant hedge ratio, $\gamma_t \sum_i^N \log(S_t^i)$ when the rolling OLS regression is applied, and $\sum_i^N \gamma_t^i \log(S_t^i)$ if the Kalman Filter algorithm applied. The performance of strategy evaluated using the following metrics: 1) Cumulative return, 2) Annualized Return, 3) Sharpe Ratio, 4) Maximum drawdown, 5) beta.

5. Back Testing-Performance Evaluation

Five different time periods was studied. In each case one year data collection was used in order to make the entire test and construct the synthetic asset. After that the algorithm starts trading with ending day for all cases the 30/12/2015. All the sample periods started at the first day of the year and ending at the last of the same year. The trading started on the first day the next year. In **Table 1** the back testing periods are presented. In **Figure 2** and **Figure 3** the results of cumulative return of Multivariate Kalman Filter algorithm against S&P500 and its maximum drawdown are shown. In **Figure 4** and **Figure 5** the cumulative return of rolling OLS regression algorithm against S&P500 and the cumulative return of rolling OLS regression algorithm against constant hedge ratio are illustrated. In **Table 2** the sets of synthetic assets are presented while in **Table 3** the name of symbols is given. In **Tables 4-8** all the metrics of the algorithms are displayed.

It can be shown that the dimension of the set and the constituents are different from period to period. This is an evidence of the non constant cointegration behavior of the stocks with the index but as we can see from the graphs the synthetic asset of each period continues to trade with profit and good metrics results.

Table 1. Back testing periods.

2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
In Sample	Trading								
	2007	2008	2009	2010	2011	2012	2013	2014	2015
	In Sample	Trading							
		2008	2009	2010	2011	2012	2013	2014	2015
		In Sample	Trading						
			2009	2010	2011	2012	2013	2014	2015
			In Sample	Trading					
				2010	2011	2012	2013	2014	2015
				In Sample	Trading				

Table 2. Sets of stock in synthetic asset on each period.

Trading Period	Stock of Synthetic Asset (Set S)	Dim (S)
2007-2015	{BRK.B, DVN, IBM, SPG}	4
2008-2015	{CAN, BAX, BMY, DELL, EMC, GE, HAL, MDLZ, NSC, T, TXN, VZ}	12
2009-2015	{DVN, FCX, IBM, MON, SPG}	5
2010-2015	{APA, AXP, C, CAT, COF, FCX, IBM, MMM}	8
2011-2015	{COST, FDX, GS, QCOM}	4

Table 3. Union set of companies raised from cointegration test from all the period of study.

APA	APACHE CORP	FDX	FEDEX CORPORATION
AXP	AMERICAN EXPRESS COMPANY	GE	GENERAL ELECTRIC CO
BAX	BAXTER INTERNATIONAL INC	GS	GOLDMAN SACHS GROUP INC
BMY	RISTOL MYERS SQUIBB COMPANY	HAL	HALLIBURTON CO (HOLDING CO)
BRK.B	BERKSHIRE HATHWAY INC (HLDG CO) B	IBM	INTL BUSINESS MACHINES CORP
C	CITIGROUP	MDLZ	MONDELEZ INTERNATIONAL INC
CAT	CATERPILLAR INC	MMM	3M COMPANY
COF	CAPITAL ONE FINANCIAL CORP	MON	MONSANTO COMPANY
COST	COSTCO WHOLESALE CORP	NSC	NORFOLK SOUTHERN CORP
DELL	DELL INC	QCOM	QUALCOMM INC
DVN	DEVON ENERGY CORP (NEW)	SPG	SIMON PROPERTIES GROUP INC
EMC	EMC CORPORATION	T	AT&T INC. COM
FCX	FREEPORT-MCMORAN COPPER&GOLD B	TXN	EXAS INSTRUMENTS INC
		VZ	VERIZON COMMUNICATIONS

Table 4. Statistical metrics performance for the trading period 01/01/2007-30/12/2015.

In Sample Period		01/01/2006-31/12/2006		
Out of Sample Period		01/01/2007-30/12/2015		
		Algorithm		
		MultiVariate Kalman Filter	Rolling (OLS) regression	Constant Hedge Ratio
Cumulative Return %	387.069	24.637	12.678	45.592
Annual Return %	19.261	2.481	1.337	4.268
Sharpe Ratio	3.298	0.892	0.483	0.303
Beta	0.019	0.017	0.027	
Maximum Drawdown %	12.642	2.925	5.666	62.453
Duration of Maximum Drawdown	73.0	147.0	390.0	1365.0

Table 5. Statistical metrics performance for the trading period 01/01/2008-30/12/2015

In Sample Period	01/01/2007-31/12/2007			
Out of Sample Period	01/01/2008-30/12/2015			
	Algorithm			
	<i>Multi Variate Kalman Filter</i>	<i>Rolling (OLS) regression</i>	<i>Constant Hedge Ratio</i>	<i>S&P500</i>
Cumulative Return %	171.671	22.347	6.855	42.470
Annual Return %	13.321	2.556	0.833	4.528
Sharpe Ratio	1.135	0.738	0.228	0.312
Beta	−0.010	−0.005	−0.013	
Maximum Drawdown %	28.636	6.730	5.301	52.909
Duration of Maximum Drawdown	175.0	404.0	628.0	154.0

Table 6. Statistical metrics performance for the trading period 01/01/2010-30/12/2015

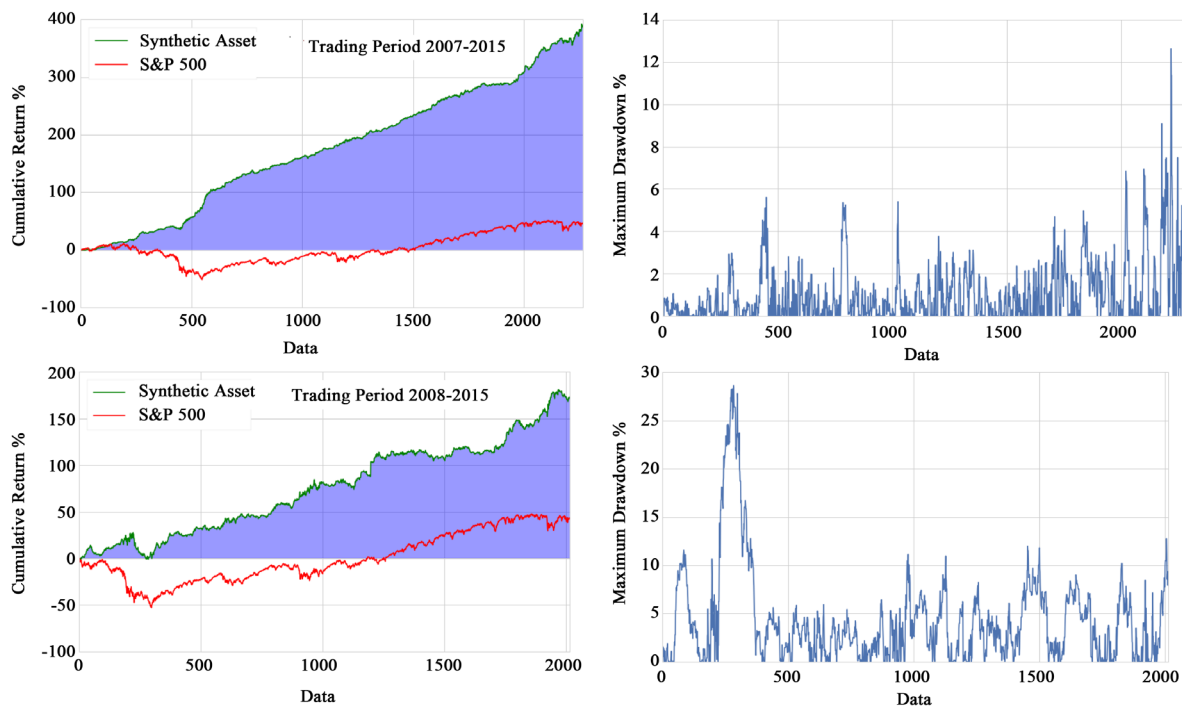
In Sample Period	01/01/2008-31/12/2008			
Out of Sample Period	01/01/2009-30/12/2015			
	Algorithm			
	<i>MultiVariate Kalman Filter</i>	<i>Rolling (OLS) regression</i>	<i>Constant Hedge Ratio</i>	<i>S&P500</i>
Cumulative Return %	278.267	28.513	13.565	121.512
Annual Return %	20.972	3.655	1.837	12.054
Sharpe Ratio	3.039	1.387	0.753	0.729
Beta	0.026	−0.003	0.032	
Maximum Drawdown %	17.910	3.934	3.736	28.547
Duration of Maximum Drawdown	69.000	290.000	367.000	203.0

Table 7. Statistical metrics performance for the trading period 01/01/2010-30/12/2015

In Sample Period	01/01/2009-31/12/2009			
Out of Sample Period	01/01/2010-30/12/2015			
	Algorithm			
	<i>Multi Variate Kalman Filter</i>	<i>Rolling (OLS) regression</i>	<i>Constant Hedge Ratio</i>	<i>S&P500</i>
Cumulative Return %	111.149	18.375	9.123	81.733
Annual Return %	13.294	2.857	1.469	10.491
Sharpe Ratio	2.313	1.154	0.617	0.709
Beta	−0.007	−0.015	0.060	
Maximum Drawdown %	10.025	5.959	5.145	23.420
Duration of Maximum Drawdown	223.0	393.0	770.0	203.0

Table 8. Statistical metrics performance for the trading period 01/01/2011-30/12/2015

In Sample Period	01/01/2010-31/12/2010			
Out of Sample Period	01/01/2011-30/12/2015			
	Algorithm			
	<i>MultiVariate Kalman Filter</i>	<i>Rolling (OLS) regression</i>	<i>Constant Hedge Ratio</i>	<i>S&P500</i>
Cumulative Return %	81.346	14.694	5.063	62.094
Annual Return %	12.674	2.787	0.995	10.167
Sharpe Ratio	2.643	1.408	0.532	0.706
Beta	-0.003	-0.002	0.040	
Maximum Drawdown %	8.515	2.070	3.373	20.889
Duration of Maximum Drawdown	125.0	143.0	255.0	203.0

**Figure 2.** Cumulative return and drawdown diagram for multivariate kalman filter algorithm for the trading periods starting at 2007, 2008 and ending at 2015.

Comparing the metrics of each period of study it is clear the Multivariate Kalman Filter algorithm gives the better results among the other and the Index S&P500. In the first period where the Index presents the higher Maximum Drawdown equal to 62% and cumulative returns (CR) equal to 45.59% (Annualized percentage APR = 4.26%) the Multivariate Kalman filter algorithm (MKFA) gives the highest cumulative rate result = 387% (with APR = 19.26%) with a reasonable Maximum Drawdown (MDD) = 12.64 and duration = 73 days. The MKFA in other cases shows max CR = 278% (APR = 20.9%) and Sharpe Ratio = 3.039 while in trading period 2008-2015 present the higher MDD = 28.63 with duration = 175 days. At the same period the Index presents MDD = 52.9% with duration = 154 days. The profit from MKFA given by CR = 172% (APR = 13.32) and on the other hand S&P500 has CR = 42.47% (APR = 4.52%). In this period the MKFA has the lowest SR = 1.13 but still higher than Index. The lower CR of MKFA given in the period 2011-2015 and equal to 81.34% (APR = 12.67%) with MDD = 8.51% and duration = 125 days. In this trading period the index has CR = 62% (APR = 10.16%) but

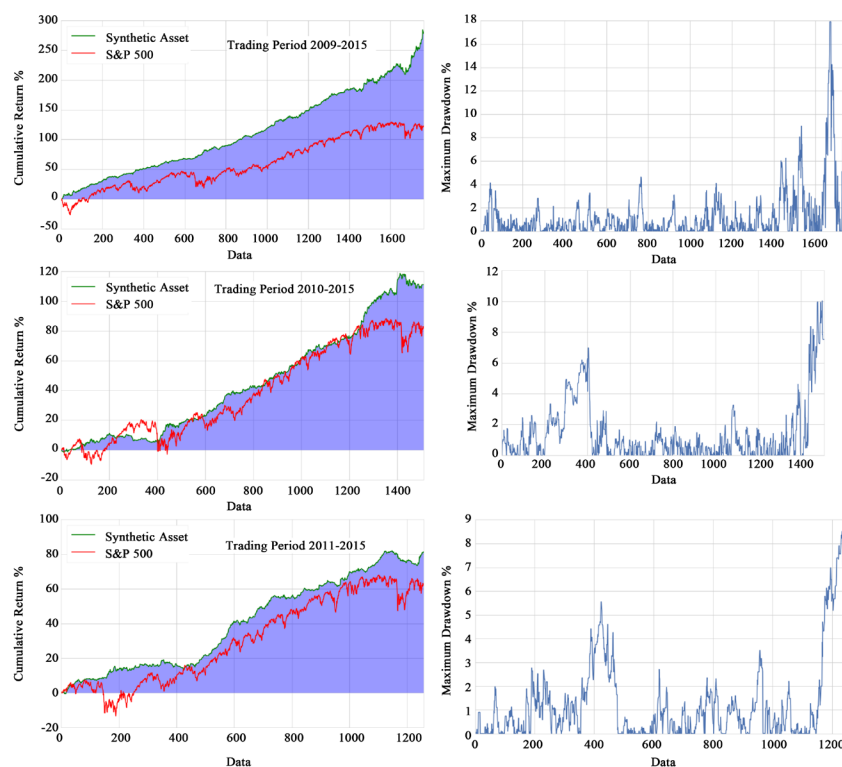


Figure 3. Cumulative return and drawdown diagram for multivariate kalman filter algorithm for the trading periods starting at 2009, 2010, 2011 and ending at 2015.

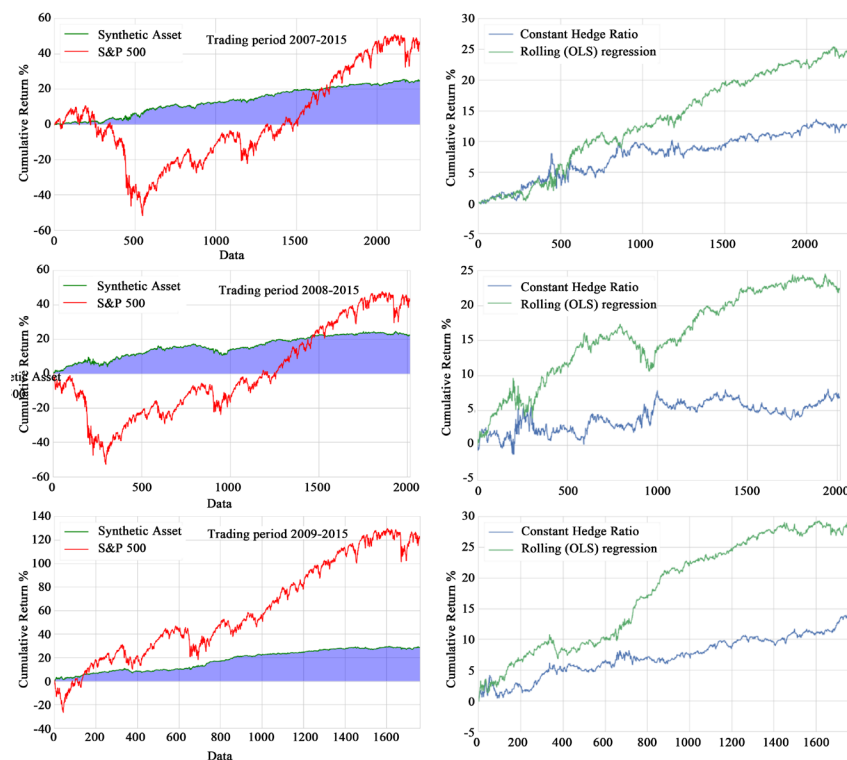


Figure 4. Cumulative return of rolling ols regression algorithm against S&P500 and algorithm with constant hedge ratio, for trading period starting AT 2007, 2008, 2009 and ending at 2015.

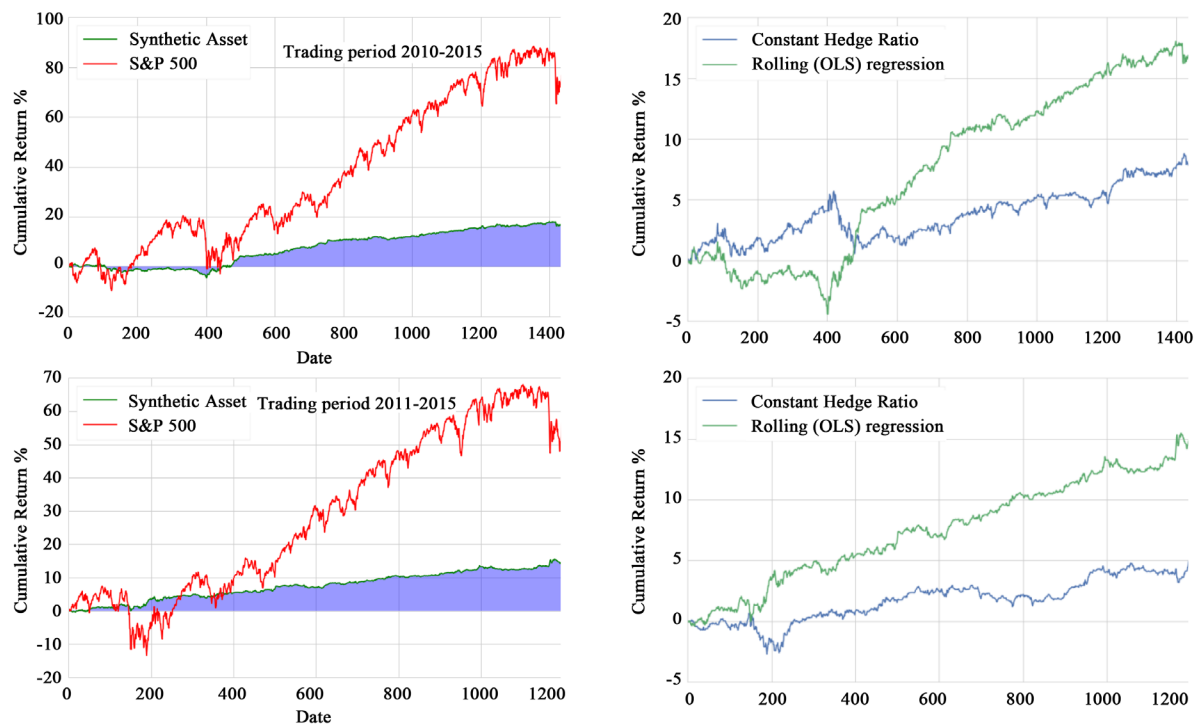


Figure 5. Cumulative return of rolling OLS regression algorithm against S&P500 and algorithm with constant hedge ratio, for trading period starting at 2010, 2011 and ending at 2015.

with higher MDD = 23.42 and duration = 203. The SR of MKFA is still higher and equal to 2.313. In the last periods the profits declined but kept higher profit and Sharp Ratio than S&P500.

The second algorithm of rolling OLS regression gives lower cumulative profit than Index but has almost doubled Sharp Ratio than S&P500 in all cases. Finally the evaluation of metrics has shown that the MKFA can beat the market as it creates statistical arbitrage condition in Index.

6. Conclusion

Mean-reverting algorithms with time adaptive hedge ratio are presented. A methodology of a synthetic asset construction based on the stocks of S&P500 has been discussed. The criterion of the selection was the cointegration relationship of individual stocks as the logarithmic sum of them with S&P500. The results of back testing show that for different period of study the form and the dimension of the synthetic asset are different. Pair trading strategy was adopted and the evaluation of the metrics results presented better behavior of MKFA among the others and beat the market. In the last periods the profits declined but it was still higher than S&P500 with much higher Sharp Ratio. The algorithm defended better its profit as the Maximum Draw down was quite lower than Index.

References

- [1] Engle, R. and Granger, C. (1987) Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, **55**, 251-276. <http://dx.doi.org/10.2307/1913236>
- [2] Vidyamurthy, G. (2004) Pairs Trading, Quantitative Methods and Analysis. John Wiley & Sons, Hoboken.
- [3] Infantino, L. and Itzhaki, S. (2010) Developing High-Frequency Equities Trading Models. Master of Business Administration, Massachusetts Institute of Technology, Cambridge.
- [4] Ernie, C. (2013) Algorithmic Trading: Winning Strategies and Their Rationale. John Wiley & Sons, Hoboken.
- [5] Caldeira, J. and Moura, G.V. (2013) Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. <http://dx.doi.org/10.2139/ssrn.2196391>

-
- [6] Dunis, C.L. and Ho, R. (2005) Cointegration Portfolios of European Equities for Index Tracking and Market Neutral Strategies. *Journal of Asset Management*, **6**, 33-52. <http://dx.doi.org/10.1057/palgrave.jam.2240164>
 - [7] Dunis, C.L., Giorgioni, G., Laws, J. and Rudy, J. (2010) Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities. CIBEF Working Papers, CIBEF.
 - [8] Elliott, R., van der Hoek, J. and Malcolm, W. (2005) Pairs Trading. *Quantitative Finance*, **5**, 271-276. <http://dx.doi.org/10.1080/14697680500149370>
 - [9] Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (2006) Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies*, **19**, 797-827. <http://dx.doi.org/10.1093/rfs/hhj020>
 - [10] Khandani, A. and Lo, A.W. (2007) What Happened to the Quants in August 2007. <http://web.mit.edu/Alo/www/Papers/august07.pdf>
 - [11] Rajamani, M. (2007) Data-Based Techniques to Improve State Estimation in Model Predictive Control. PhD Thesis, University of Wisconsin-Madison, Madison.
 - [12] Rajamani, M.R. and Rawlings, J.B. (2009) Estimation of the Disturbance Structure from Data Using Semidefinite Programming and Optimal Weighting. *Automatica*, **45**, 142-148. <http://dx.doi.org/10.1016/j.automatica.2008.05.032>