

Received: 2017-10-14

Accepted: 2017-12-21

Online published: 2018-01-10

DOI: <https://doi.org/10.15414/meraa.2017.03.02.85-94>*Original Paper*

Statistical arbitrage regarding trade in goods of agricultural origin

Vladimír Matušek^{1*}, Maciej Sporysz², Krzysztof Molenda²

¹ Slovak University of Agriculture, Faculty of Economics and Management, Department of Mathematics, Nitra, Slovak Republic

² Center of Computation, New Technologies and Informatics, University of Agriculture in Krakow, Kraków, Poland

ABSTRACT

In our work we concentrate on the use of statistical arbitrage, mainly the trade in goods of agricultural origin model, which considers high frequency trading mechanisms of goods of agricultural origin. The strategy usefulness was measured by the information ratio. We considered both, options in the horizon and the time interval. The best results were achieved with respect to the historical time horizon, where the benchmark proportion of securities and one-minute time interval were appointed. Although we have to admit the investment issue constitutes a very complex problem influenced by a large number of factors. So there is not any universal mode of conduct guaranteeing profits that may be unequivocally indicated. We can only define a scenario, which will be effective with a substantial degree of probability.

KEYWORDS: statistical arbitrage, frequency trading, Kalman filter, simulation, investment strategy

JEL CLASSIFICATION: Q02, Q13, Q17

INTRODUCTION

Algorithmic trading stems from uncomplicated applications, which make it possible to divide large orders into a few smaller ones and realize them in an optimal manner. The development of this technique was possible only when the on-line stock exchange market allowing sending orders by e-mail became common and when personal computers became available. The current programmes use complex algorithms containing mathematical tools, especially in the field of statistics, optimisation or probability theory, presented Esfahanipour and Mousavi [5]. By observing quotations and other sources having a direct influence on the stock exchange markets, transaction programmes give adequate investment signals. As a result of the robust development of the IT services and an easy access to the Internet, the access to algorithmic trading is not reserved for large and significant investors anymore, but it is also available for individual investors. Enterprises, which provide such software, compete on many levels. First of all, they endeavour to generate highest possible profits for the customer (consequently, they

* Vladimír Matušek, Slovak University of Agriculture, Faculty of Economics and Management, Department of Mathematics, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic. E-mail: vladimir.matussek@uniag.sk

increase their nominal commission by increasing the number of licenses and the amount of capital). Secondly, they try to indicate incorrect conduct of robots applied by the competition. Finally, and most importantly, they modify the operation of their algorithms with regard to moves and strategy adopted by alternative algorithms, was presented by Chavarnakul and Enke [3]. Thus, one may very rarely find in literature details concerning algorithms used by robots. And even if one succeeds, these are mainly publications relating to the application of artificial intelligence, the co-called 'black boxes'.

Statistical arbitrage trading has previously been examined by various authors, Dunis and other [4]. The goal of this type of trading is to develop highly automated trading strategies. Bertram [1] presented analytic formulae and solutions for calculating optimal statistical arbitrage strategies with transaction costs. The author assumed that the traded security had been described by an Ornstein–Uhlenbeck process. Broussard and Vaihekoski [2] investigated the practical issues of implementing the self-financing pairs portfolio trading strategy. These strategies engage in high frequency trading using algorithms based on stochastic or deterministic methods to identify price inefficiencies in the market [6]. A common approach when performing this type of trading is to construct a stationary, mean reverting synthetic asset.

MATERIAL AND METHODS

The Kalman filter has been used in order to forecast the cointegration coefficient in an optimal way. In the 1960's Kalman introduced the recursive estimation algorithm using the least squares method of parameters variable in time. Its fundamental idea is based on the predictor-corrector system and shown in the flow chart presented in Figure 1.

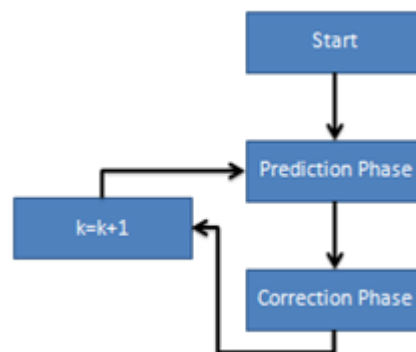


Figure 1 The idea of the Kalman filter operation. Source: authors

The a priori condition vector is predicated during the first phase.

$$\beta_{(t|t-1)} = \beta_t$$

Then it undergoes correction.

$$v_t = Y_t - X_t \beta_t$$

$$F_t = X_t P_t X_t' + H$$

$$\beta_{t+1} = \beta_t + P_t X_t' \frac{v_t}{F_t}$$

Finally, the covariance matrix is recalculated:

$$P_{t+1} = P_t + P_t X'_t X_t P_t \frac{1}{F_t} + Q$$

The H and Q parameters used in the procedure are responsible for smoothing time errors. The F matrix is the so-called state transition matrix.

Pair trading

The initial value is taken as

$$\beta_1 = \frac{Y_1}{X_1}.$$

The only condition of the adopted strategy was that the pair of commodities was supposed to originate from the same economic sector. The idea of pair trading itself is based on two pillars: the purchase of a commodity, whose price is determined as too low and a short sale of a commodity, whose price is too high, according to the investment algorithm. Thus, the strategy is sometimes called the hedging strategy. The future direction of price changes of the respective goods is of no importance, i.e. it is not essential whether the price of a single commodity shall increase or decrease. What is important is the direction of the changes of the price ratio between both of them.

Let's assume that we have at our disposal two time series (P_{Y_t} and P_{X_t}) of the prices of selected goods. Moreover, let's suppose that they are cointegrated with each other. With the help of the Kalman filter we may calculate the forecast price of the first of them as

$$P_{Y_t}^* = \beta_t P_{X_t}$$

and the difference in prices

$$z_t = P_{Y_t} - P_{Y_t}^* = P_{Y_t} - \beta_t P_{X_t}.$$

The time series determined this way undergoes normalisation by cutting the average and dividing by the standard deviation. It should be stressed that the procedure of determining the average and the standard deviation is applied only with respect to the historical sample (in-sample), these data are used only to establish the constant proportion of goods prices. The out-of-sample specimen also undergoes normalisation, but on the basis of the historical average and the standard deviation. Here we may already make investment decisions.

Issues of optimizing the threshold of position opening and closing have not been considered in the study. Therefore, for the purposes of this work it has been assumed that exceeding the 'spread' i.e. the difference in price and its estimation, regarding the absolute value by the standard deviation value, constitutes the level of position opening (purchase of one commodity and short sale of the other, where ± 1 signal mark decides which commodity is purchased and which is subject to short sale). Furthermore, it has been assumed that the actual position opening should occur only following two successive signals in order not to have the position opened too early. The position closing (sale of one commodity and purchase and return to the broker of the other commodity) is only possible when the difference in prices decreases below half of the standard deviation value.

Data collection method

Data concerning quotations of selected securities have been collected for the period of one year from April 1st 2013 to March 31st 2014 with one minute interval.

Statistical measures of investment strategy evaluation

The most common investment ratios have been applied for the purpose of evaluating quality of the investment strategy using the statistical arbitrage in high frequency trading.

Annualised return

This ratio is applied when the strategy testing period is different from the period of one year. It allows adjusting the rate of return from any period to one-year time horizon. Owing to this, it is possible to compare strategies of different time horizons

$$RA = 252 \cdot \frac{1}{N} \sum_{t=1}^N R_t,$$

where R_t - daily return on investment, N - number of days under analysis.

Annualised volatility

The annualised volatility ratio is a measure of statistical return dispersion for a given security or market index. It may be measured with the help of the standard deviation or variation between revenues derived from the same security or market index. It shows the extent of risk and uncertainty, which may occur on the stock exchange

$$\sigma_A = \sqrt{\frac{252}{N-1} \cdot \sum_{t=1}^N (R_t - \bar{R})^2},$$

where:

R_t - daily return on investment,

\bar{R} - average daily return on investment,

N - number of days under analysis.

Information ratio

The Information ratio (IR) is one of the most common indexes used for the purpose of comparing the risk level of various investment strategies. It is a measure of results achieved by the administrator with respect to the adopted risk level

$$IR = \frac{RA}{\sigma_A}.$$

It may be expressed in negative and positive values. The positive values are the most beneficial. The IR values within the 0,50 – 0,75 range are regarded a good investment and within the 0,75 – 1 range - a very good investment. The coefficient values exceeding 1 are evidence of an unusually good investment.

Maximum drawdown

The maximum drawdown ratio allows estimating the maximum loss that might flow from the investment. It describes the worst possible investment scenario, i.e. purchase at maximum prices and sale at minimum prices within the examined period:

$$MD = \min_{i=1, \dots, t; t=1, \dots, N} \sum_{j=1}^t R_j$$

Implementation of statistical arbitrage in Matlab 2014 environment

The prototype of the statistical arbitrage consists of ten modules. Start constitutes the main module, which steers the realisation succession and the communication between the respective modules. The flow chart of component realisation succession of the prototype is shown in Figure 2.

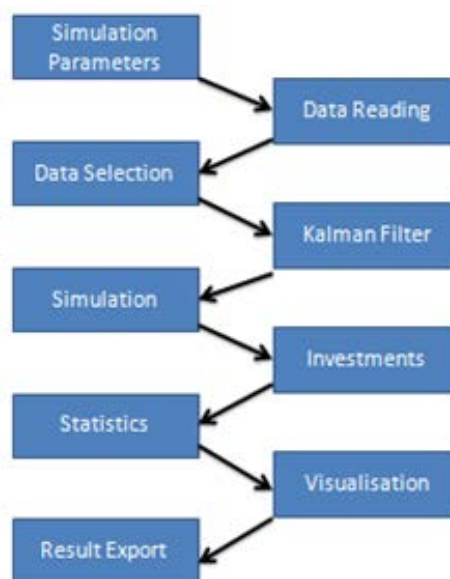


Figure 2 Scheme of communication between respective modules of the prototype. Source: authors

RESULTS AND DISCUSSION

Determination of simulation parameters is the module responsible for determining simulation parameters. The names of variables and their possible values and description are demonstrated in Table 1.

The objective of this module is to determine statistics concerning the investment profitability. The return on investment, the basic parameter, is calculated in the following way. The **Annualised return** (Annualised_Return) is the second parameter, which allows determining the profitability, but this time on a yearly basis. Determining **volatility** of the expected **Annualised return** (Annualised_Volatility), and consequently, the risk the investor should take into account, conducting such transactions, is the next factor. Furthermore, the **IR** (Information ratio) coefficient has been determined as the ratio of the annualised return to the annualised volatility. The **Maximum drawdown** (MaxDrawDown) has also been defined. Figure 3 demonstrates the total pecuniary value of available funds and means invested in commodities within the successive time moments of quotations. This diagram derives from the prototype of the statistical arbitrage strategy in High Frequency Trading implemented in the Matlab environment.

Table 1 Simulation parameters

Variable	Possible Value	Description
startDate	'03/03/2013'	The initial date from which the quotation analysis begins. This solution allows selecting from long-term historical data only the period concerned with respect to the analysis, without the necessity to additionally process the data.
startTime	'9:30'	Listing initial time. Records including pre-session transactions may be found in the historical data.
endTime	'16:00'	Listing finishing time. Records including post-session transactions may be found in the historical data.
closeTime	'15:45'	Determination of time, after which transactions are not performed, despite the signal. The risk of unfavourable rate fluctuations between the next session days is limited then. \ Entering data into sale transactions is possible only between <i>closeTime</i> and <i>endTime</i> .
endTime	100	Number of days of simulation duration.
N	2	Number of days included in the history. The parameter of simulation memory.
Interval	5	Time interval, within which transactions may be entered into. In this case the change of the security number may be conducted every 5 minutes.
K	4	The procedure regarding temporary quotation suspension has been included in the simulation. In such cases data gaps are estimated with the help of the weighted average from the last quotations.
stopLoss	20	The parameter expressed in percentage, which prevents the phenomenon of dip on the stock exchange. In case of the sudden drawdown the possessed securities become sold out.
Capital	10000	The original value of funds intended for investments.
Times	100	The multiple of stock number (the number of stocks in the block).
brokerCost	0.0005	Broker's commission - amount per each block of stock.
bidaskCost	0.0005	Commission per transaction.
oSTreshold	-1.0	The threshold, below which the security purchase signal is generated (the market value is underestimated).
oLTreshold	1.0	The threshold, above which the security sale signal is generated (the market value is overestimated).
cTreshold	0.5	The threshold, at which the signal was generated earlier and remains binding.
H	0.01	The parameter used with the Kalman filter.
NoiseRatio	0.0001	The parameter used with the Kalman filter.

Source: authors

It may be noticed that the strategy after two days of operation has resulted in the profits at the level of 1.2 %. It is worth emphasizing that the strategy does not always guarantee profits. The investor should also take into account the possibility of suffering losses, which is reflected in the Figure 3 above.

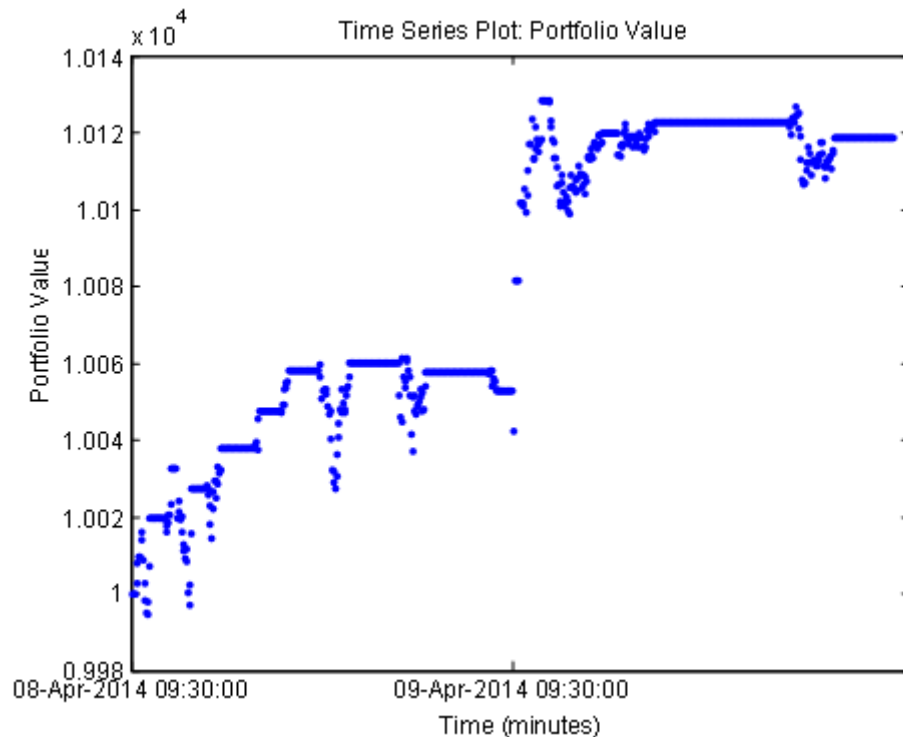


Figure 3 The portfolio value depending on listing time (the 1-minute time interval)

Table 2 demonstrates the basic statistics evaluating its quality.

Table 2 Statistics summarising the strategy simulation

Ratio	Value
Annualised Return	149.64%
Annualised Volatility	9.46%
Information Ratio	1.58
MaxDrawDown	\$9.70

Source: authors

One should, first of all, concentrate on the high annualised return. Assuming the strategy operation analogical to the operation presented earlier and the capital amounting to \$10,000.00, the investor could count on the profit amounting to \$14,900.00. The IR ratio above 1, reflecting a very good investment, serves as the confirmation of quality. The maximum drawdown at the level below \$10.00 should not deter the investor, taking into consideration the invested capital. In addition to this, the risk connected with the investment is significantly lower than expected profits.

The next step of the analysis is to check if the time interval length influences the strategy quality. Figure 4 show that the shorter time interval does not always mean a higher return on investment.

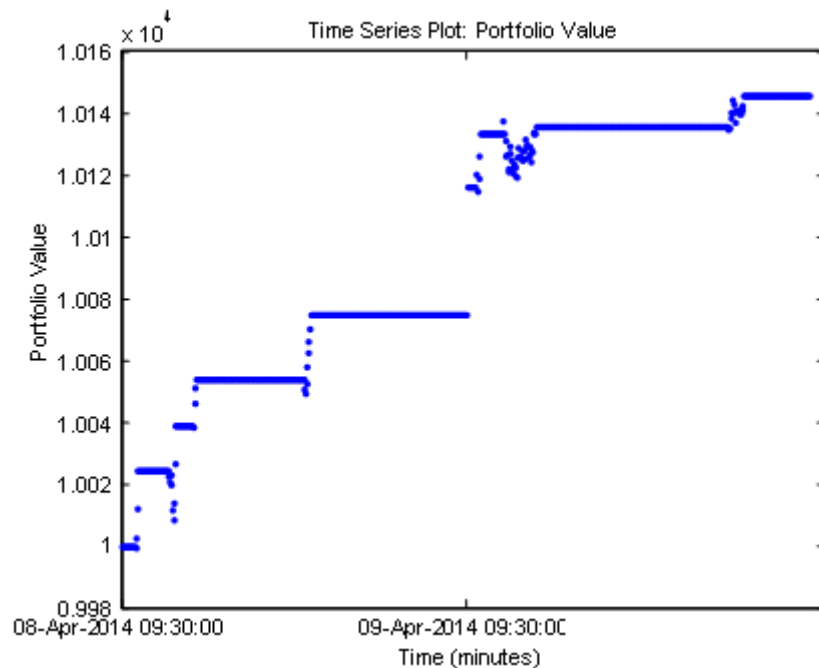


Figure 4 The portfolio value depending on listing time (the 5-minute time interval) Source: authors

Due to the fact that Figure 4 is for reference only (they are based on one arbitrarily chosen day), the researchers have tried to verify the influence of the time interval on the strategy quality. For this purpose the time horizon of simulation duration of 100 days has been adopted in the analysis, whereas 1, 5, 10 and 15 minutes respectively have been adopted as the interval variable. Other input parameters are shown in Table 3.

Table 3 Simulation Input Parameters

Parameter	Value
startDate	01-04-2013
start time	09:30
endTime	16:00
closeTime	15:45
N	1
sym_len	100
Capital	10000
Times	10
oSTreshold	-1
oLTreshold	1
cTreshold	0.5

Source: authors

The results are presented in Table 4. The highest rate of return have been achieved in case of the 15-minute time interval (over 20 % of return within a year), however, the result is burdened with a considerable risk (64 % annualised volatility). A slightly lower annualised return (nearly 18 %) has been recorded for the 1-minute interval, but the result has been more

beneficial with respect to the investment risk (55 % annualised volatility). A significantly lower annualised return has been recorded for the interval (6 %). The relatively low risk suffered (25 %) argues in favour of this option. The 5-minute time interval, where a loss over 2 % has been recorded, has turned out to be the worst of the subject options. None of these investments has turned out to be good ($IR < 0.5$), although the 1-, 10- and 15-minute intervals have undoubtedly been favourable. Despite the fact that the 10-minute option has resulted in a small profit, it has been characterised by the maximum drawdown. Interestingly, this coefficient is comparable in terms of the most favourable and the worst annualised return.

Table 4 Summarising statistics depending on the time interval

Time interval [min]	Ratio			
	Annualised Return [%]	Annualised Volatility [%]	Information Ratio []	Maximum DrawDown [\$]
1	17.81	55.12	0.32	384.69
5	-2.32	34.65	-0.06	465.54
10	6.46	25.29	0.25	523.65
15	20.19	64.02	0.31	479.81

The other parameters have not been modified, just as in Table 5. It may be observed that owing to the increase in the number of days considered for determining the proportion model, the increase in the annualised return has been recorded, where the sample with the 5-day time horizon has been the highest (42 %). However, a high level of risk (44 %) has been recorded here. It is worth emphasizing that lengthening the time horizon has not always translated into the increase in the annualised return. Each investment with the horizon longer than one day has been evaluated either as a very good investment ($IR [0.75; 1)$) or an unusually good investment ($IR > 1$).

Table 5 Summarising statistics depending on the number of days in the history

Number of days in the history	Ratio			
	Annualised Return [%]	Annualised Volatility [%]	Information Ratio []	Maximum DrawDown [\$]
1	17.81	55.12	0.32	384.69
2	30.34	24.84	1.22	177.52
3	31.35	22.13	1.41	45.11
4	25.18	24.34	1.03	59.88
5	41.75	43.88	0.95	59.12

Source: authors

CONCLUSIONS

In view of the above arguments, one may not clearly state which time interval is the most favourable. Monitoring historical data of several time interval options may constitute the solution of this problem, whereas adopting the time interval which has recently turned out to be the best one may serve for the purpose of investing. Formulation of a periodical mechanism of the time option change might be proposed as a constructive solution. The

model of the statistical arbitrage with high frequency trading in goods of agricultural origin may constitute an interesting alternative for traditional selling methods.

REFERENCES

- [1] Bertram, W. (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica A - Statistical Mechanics and its Applications*, 389(11), 2234-2243.
- [2] Broussard, J.P. & Vaihekoski, M. (2012). Profitability of pairs trading strategy in an illiquid market with multiple share classes. *Journal of International Financial Markets, Institutions and Money* 22(5), 1188-1201.
- [3] Chavarnakul, T. & Enke, D. (2008). Intelligent technical analysis based equivolume charting for stock trading using neural networks. *Expert Systems with Applications* 34(2), 1004-1017.
- [4] Dunis, Ch., Giorgioni, G., Laws, J. & Rudy, J. (2010). *Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities*. doi: <http://dx.doi.org/10.2139/ssrn.2272605>
- [5] Esfahanipour, A. & Mousavi, S. (2011). A genetic programming model to generate risk-adjusted technical trading rules in stock markets. *Expert Systems with Applications*, 38(7), 8438-8445. doi: <https://doi.org/10.1016/j.eswa.2011.01.039>
- [6] Gradojevic, N. & Gençay, R. (2013). Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking & Finance*, 37(2), 578-586. doi: <https://doi.org/10.1016/j.jbankfin.2012.09.012>