IMPROVING PAIRS TRADING WITH RESPONSE SURFACE METHODOLOGY

AN APPLICATION OF ROGER S. KUO'S STUDY ON "0001.HK CHEUNG KONG/0013.HK HUTCHISON" TO DETERMINE THE BEST COMBINATION OF SELECTION PERIOD AND TRADING PERIOD

BY

WONG HO KA
03006433
FINANCE

An Honours Degree Project Submitted to the
School of Business in Partial Fulfillment of the
Graduation Requirement for the Degree of
Bachelor of Business Administration (honours)

Hong Kong Baptist University
Hong Kong

April 2006

1. Acknowledgement

I am indebted to Dr. Joseph Fung for his supervision and guidance, especially on inspiring me to go into this topic and actually discussed critical issues in this paper. He gave me a great help hand so that I can finish this paper in limited period of time. Thank you very much to your help.

I would like to thanks those of you who support me throughout this tough semester as well.

2. Abstract

After Roger mentioned the optimized outcome by Response Surface Methodology in pairs trading, one may wonder if the same works given he/she has a favorite pair, but also limited money. Especially, can we apply the same model in Hong Kong market.

In this study, as an application of Roger' s work to Hong Kong market, we are going to test the model only with time being the independent variable, so as to find out the time effect on pairs trading. If the result is positive, we also wonder what the best strategy is. And also we are going to see if this method works on individual pairs instead of the entire pool of pairs.

The result shows that Response Surface Methodology failed to optimize the return of selected pair. But it inspires us that to the selected pair, time is a favorable and significant factor to the trading result. In future, we can add more variables so that we can actually test if this method gives us the best combination of time.

3. Introduction

3.1. Background

Roger has demonstrated us the importance of determining selection period, trading period, threshold and number of pairs during pairs trading. A wise choice can greatly enhance the trading result.

Roger's research had been focusing on US market, as well as it generalized the whole market's situation instead of individual pair's performance, so small investors in Hong Kong, who have limited money for only one pair, may wonder if Roger's research helps them. "0001.HK Cheung Kong" and "0013.HK Hutchison" had been chosen to be the testing pair. Though it may not project the market situation, as long as it is positive, that means small investors had one more choice and future researches can be done to figure out the best strategies for various pairs.

This paper is going to follow the framework set by Roger, with some modifications to Hong Kong market, so that we can investigate the possibility of using Response Surface Methodology to improve our trading results by only varying the period of time. Besides, the paper focus on the effect of varying selection period and trading period, which are the most foundation elements of pairs trading.

3.2. Literature Review

3.2.1. According to Roger, the best combination in US market is[1]:

Selection Period	10.17 Months
Trading Period	9 Months
Number of Pairs	3
Divergence Threshold	1.5 Standard Deviation
Shorting Margin	0.5

3.2.2. The Experimental Setup is designed first to test the significance of each factor's effect, so as to determine its presence in next model. It is followed by Response Surface Model, which optimized the input factors. Details will be explained in part 5.

3.2.3. A computer program is written to divide the data. The first q days are used to analyze the population mean and divergence threshold. The following p days are the period we actually trade. Then the software will generate the return of each run *n*.

3.3. Statement of the problems

3.3.1. To investigate whether it is feasible to choose the best selection period and trading period by Response Surface Methodology during pairs trading so as to obtain the best trading result.

[1] Roger S. Kuo, (2004), Improving Pairs Trading with Response Surface Methodology, p.18

3.4. Objectives

This paper seeks a way to improve the profitability of pairs trading, even with a small initial investment. Time factor is the major concern in this paper. Since for small investors, the most flexible but less costly factor to them is time.

The other objective is to investigate the problems raised, if any, with the Response Surface Methodology so we can improve the strategy in future.

3.5. General Assumption

3.5.1. Historical data can predict future price movement.

3.5.2. There are 22 trading days per months and 264 trading days a year.

3.5.3. We can sell a fraction of one share.

3.5.4. The broker has enough shares of our selected pair so we can borrow anytime we want.

3.5.5. There is no short squeeze

3.5.6. Assume Roger is right that 1.5 standard deviations is the optimal threshold to small investors, that they cannot cover the short margin if the threshold is too high.

3.5.7. A pair is converge when the price ratio pass through the mean.

4. Data

4.1. Daily adjusted close prices of "0001.HK Cheung Kong" and "0013.HK Hutchison" dated from 1/2/2004 to 12/30/2005 are obtained from DataStream® and used in this project. One month is considered equivalent to twenty two trading days. If 05 trading day is not sufficient, the trade will end at last day of December. One year of time is used to avoid validity risk when the correlation coefficient of the pair changes.

5. Methodology

5.1. Procedure of Trade

5.1.1. When the price ratio is within 1.5 standard deviation of mean, do nothing.

5.1.2. When the price ratio is over/under the threshold, open position; we long the underperformed and short the outperformed.

5.1.3. If the price ratio remains upper/lower than the threshold, do nothing.

5.1.4. If the price ratio pass through mean, unravel the position.

5.1.5. If trading period is up/it is the end of 2005, unravel the position.

5.2. Experimental Setup

5.2.1. The full factorial of x1 and x2 is shown above. Given +1 = 264 Days (12 Months), -1 = 132 Days (6 months), 0 = 198 Days (9 months). x_1 is selection period and x_2 is trading period.

run	x_1	x_2	Response
1	-1	-1	Y_1
2	+1	-1	Y_2
3	-1	+1	Y_3
4	+1	+1	Y_4

5.2.2. We then calculate the effects of each factor. Let $P(x_1)$ to be the total number of observation when x_1=+1, and $M(x_1)$ to be the set of run when x_1 = -1.

$$\text{Effect}(x_i) = 1/|P(x_i))|\ \text{sum}(y_p) - 1/|M(x_i)|\ \text{sum}(y_m)$$

This formula can also evaluate the effects of multi-factor interactions. For example, if we study the effects of x_1x_2, we simply multiply the composites, that is +1,-1,-1,+1 so we can obtain the effects as well.

5.2.3. The next step is to find out the significance of these two effects to the selected pair. Given m = number of observations in population and m_i = number of observations in each run, we calculate the pooled variance by taking an average of these sample variance.

Sum [(Sample Variance) x (m_i/m_i)] = Pooled Variance

Then we convert it into effect variance:

Pooled Variance x 4/m = Effect Variance

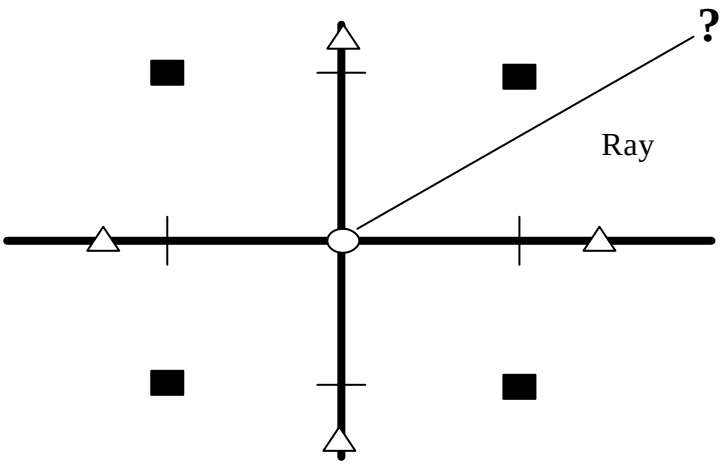
5.2.4. Then we construct an interval with t-stat value at given effect standard deviation, if it does not include zero, then the effect is statistically significant at 5% level.

5.3. Response Surface Methodology

5.3.1. Run 5 is added as central point run and run 6-9 are the axial run to aid estimating the gradient's direction.

run	x_1	x_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1
5	0	0
6	-2^(1/2)	0
7	+2^(1/2)	0
8	0	-2^(1/2)
9	0	+2^(1/2)

5.3.2. Imagine there is a ray coming out from the center point, the more surface it occupies, the better result it is.



Full Factorial Points (x_i = +1/-1) ■

Center Points (x_i = 0) ⬭

Axial Points (x_i = |2^(1/2)|) △

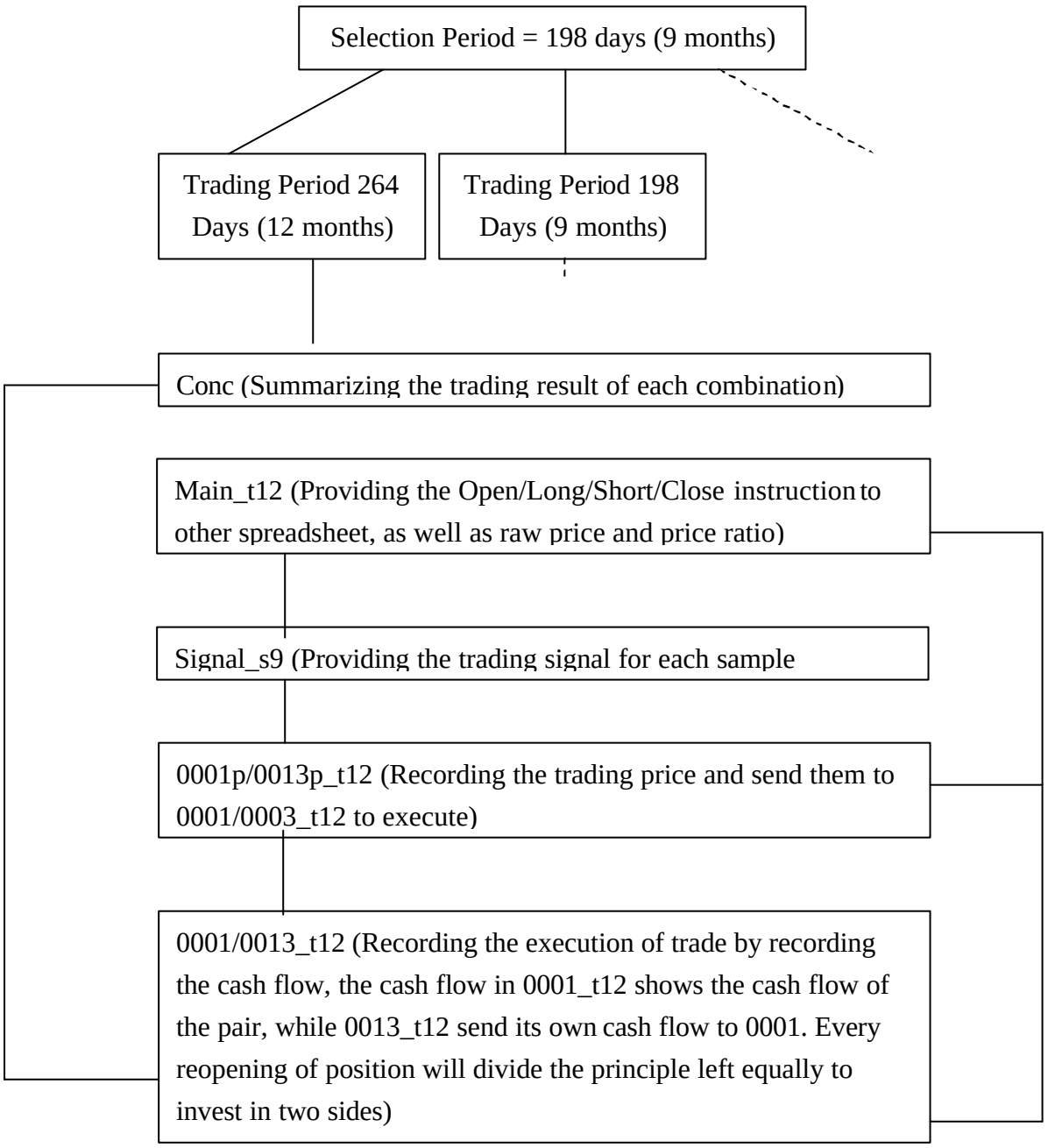
5.3.3. We can use the following regression model to estimate the ray's direction:

$$Y = C_0 + \text{Sum}(C_i X_i) + \text{Sum}\,(C_{ij} X_i X_j) + e$$

5.3.4. After obtaining the significant coefficient, we form a linear regression that let us know the direction of the ray, so we can go along its gradient to find the best combination.

5.4. Excel Workbooks

5.4.1. Instead of C language, Excel 2003 is used to analyze the data. The model is an interaction within multiple layer of worksheet and workbooks. For example:



5.4.2. Then setupresult.xls will collect all returns of different combinations of selection period and trading period. And also calculate the significance of experimental setup.

5.4.3. Tanagra 1.4.5 is used to calculate estimate the regression coefficient.

6. Results

6.1. Experimental Setup

6.1.1.

run	x_1	x_2	Response(Return in %)
1	-1	-1	-2.376709408
2	+1	-1	1.15212891
3	-1	+1	3.040380256
4	+1	+1	3.588557078
5	0	0	4.448269173

$\text{effect}(x_1) = -0.311364$

$\text{effect}(x_2) = 1.190869$

$\text{effect}(x_1x_2) = -2.678797$

UL of $\text{effect}(x_1) = -0.284145279$

LL of $\text{effect}(x_1) = -0.338583471$

UL of $\text{effect}(x_2) = 1.218087705$

LL of $\text{effect}(x_2) = 1.163649513$

UL of $\text{effect}(x_1x_2) = -2.651577562$

LL of $\text{effect}(x_1x_2) = -2.706015755$

Both of them did not include zero in the interval, indicating that they are all statistically significant at 5% level. So we should include them all in response surface methodology.

6.2. Response Surface Methodology

6.2.1. After calculating the axial points, that is 9 months plus and minus the square root of 2, we know that the new x_1 and x_2 is 167 Days and 229 Days respectively.

6.2.2.

run	x_1	x_2	Response(Return in %)
1	-1	-1	-2.376709408
2	+1	-1	1.15212891
3	-1	+1	3.040380256
4	+1	+1	3.588557078
5	0	0	4.448269173
6	-2^(1/2)	0	-0.103255291
7	+2^(1/2)	0	4.980975137
8	0	-2^(1/2)	1.555019197
9	0	+2^(1/2)	5.292567908

From Appendix 1, we know that only x_1 and x_2 is significant (greater than critical t-value at 95% confidence level), so only these two variables will be included in following steps.

6.2.3. From Appendix 1, we know that the coefficient of x_1,x_2 and the constant is 1.4084,16424,2.39755.

x_1 coef.	x_2 coef.	x_1	x_2	x_1 (Actual)	x_2 (Actual)	Constant	Response

1.4084	1.6424	0	0	9	9	2.39755	2.397548
1.4084	1.6424	0.07	0.08	9.07	9.08	2.39755	2.627528
1.4084	1.6424	0.14	0.16	9.14	9.16	2.39755	2.857509
1.4084	1.6424	0.21	0.24	9.21	9.24	2.39755	3.087489
1.4084	1.6424	0.28	0.32	9.28	9.32	2.39755	3.31747
1.4084	1.6424	0.35	0.4	9.35	9.4	2.39755	3.54745
1.4084	1.6424	0.42	0.48	9.42	9.48	2.39755	3.777431
1.4084	1.6424	0.49	0.56	9.49	9.56	2.39755	4.007411
1.4084	1.6424	0.56	0.64	9.56	9.64	2.39755	4.237392
1.4084	1.6424	0.63	0.72	9.63	9.72	2.39755	4.467372
1.4084	1.6424	0.7	0.8	9.7	9.8	2.39755	4.697353
1.4084	1.6424	0.77	0.88	9.77	9.88	2.39755	4.927333
1.4084	1.6424	0.84	0.96	9.84	9.96	2.39755	5.157314
1.4084	1.6424	0.91	1.04	9.91	10.04	2.39755	5.387294
1.4084	1.6424	0.98	1.12	9.98	10.12	2.39755	5.617274
1.4084	1.6424	1.05	1.2	10.05	10.2	2.39755	5.847255
1.4084	1.6424	1.12	1.28	10.12	10.28	2.39755	6.077235
1.4084	1.6424	1.19	1.36	10.19	10.36	2.39755	6.307216
1.4084	1.6424	1.26	1.44	10.26	10.44	2.39755	6.537196
1.4084	1.6424	1.33	1.52	10.33	10.52	2.39755	6.767177
1.4084	1.6424	1.4	1.6	10.4	10.6	2.39755	6.997157
1.4084	1.6424	1.47	1.68	10.47	10.68	2.39755	7.227138
1.4084	1.6424	1.54	1.76	10.54	10.76	2.39755	7.457118
1.4084	1.6424	1.61	1.84	10.61	10.84	2.39755	7.687099
1.4084	1.6424	1.68	1.92	10.68	10.92	3.39755	8.917079
1.4084	1.6424	1.75	2	10.75	11	4.39755	10.14706
1.4084	1.6424	1.82	2.08	10.82	11.08	5.39755	11.37704

1.4084	1.6424	1.89	2.16	10.89	11.16	6.39755	12.60702
1.4084	1.6424	1.96	2.24	10.96	11.24	7.39755	13.837
1.4084	1.6424	2.03	2.32	11.03	11.32	8.39755	15.06698
1.4084	1.6424	2.1	2.4	11.1	11.4	9.39755	16.29696
1.4084	1.6424	2.17	2.48	11.17	11.48	10.3975	17.52694
1.4084	1.6424	2.24	2.56	11.24	11.56	11.3975	18.75692
1.4084	1.6424	2.31	2.64	11.31	11.64	12.3975	19.9869
1.4084	1.6424	2.38	2.72	11.38	11.72	13.3975	21.21688
1.4084	1.6424	2.45	2.8	11.45	11.8	14.3975	22.44686
1.4084	1.6424	2.52	2.88	11.52	11.88	15.3975	23.67684
1.4084	1.6424	2.59	2.96	11.59	11.96	16.3975	24.90683

6.2.4. At the ratio of coefficient at 7:8, the maximum in this study occur at 11.59 months and 11.69 months. That is the selection period should be 255 Days and Trading Period should be 264 Days.

7. Conclusion

7.1. Obviously, the Response Surface Methodology failed to identify the maximum point. At least run 7 and run 9 is more profitable than run 4. However, the result demonstrate an interesting trend that it seems increasing the selection period and trading period can increase the profit of this selected pair in general. One reason can be the correlation coefficient of this pair is extremely stable. That means the price ratio mean and threshold seldom change dramatically. So if we use more data to analyze, we can have more precise trading signals. And if we trade in a longer period of time, we earn more by increasing number of trades. In short, the methodology tells us that the more time period it is, the better the result is. And the selection period to trading period ratio is better to be

around 7:8.

7.2. The result can also attribute to the limitation of this study. Firstly, multiple linear regression is used to estimate the response surface, whereas, the factor's relationship can be non-linear. So we may consider using non-linear regression. Secondly, there is no logic behind this study, in order words; it is merely an observation and drawing statistically conclusion. We cannot figure out if there are anything happened and suddenly alter the trend. On the other hand, in reality there are much more factors affecting the trading result. Even this study shows that increasing trading time can enhance the result of this pair; in the real world we must consider other factors like short squeeze and required margin. Thus, more variables can be tested to compose a regression that better describe the situation.

7.3. The original purpose of choosing this pair is to avoid frequent changing in correlation coefficient and thus decrease the validity. But the result shows that high level of correlation coefficient may decrease the preciseness of the model. In future, we can try more pairs which have different level of deviation in correlation coefficient to find out which one alone can be optimized by response surface methodology.

7.4. To conclude, we know that selection period and trading period is a significant factor that affects the result. Response Surface Methodology may optimize the pool of pairs in whole market. But when it is used to optimize single pair, it should be use with caution that variables and regression must be chosen properly. Next time, if we want to examine the effect of one kind of factors, like time, on a single pair, we may wish to try

another model.

8. References

i. Roger S. Kuo, (2004), Improving Pairs Trading with Response Surface Methodology

ii. Ganapathy Vidyamurthy, (2004),Pairs trading: quantitative methods and analysis, John Wiley & Sons Inc.

iii. Gerald Keller, Brian Warrack, Statistics For Management and Economics Sixth Edition, Thomson Brooks/Cole, Chapter 15

Appendix1

Global results

Endogenous attribute	**Response(Return in %)**
Examples	9
R?/TD>	0.757042
Adjusted-R?/TD>	0.611267
Sigma error	1.595696
F-Test (3,5)	5.1932 (0.053859)

Analysis of variance

Source	**xSS**	**d.f.**	**xMS**	**F**	**p-value**
Regression	39.6697	3	13.2232	5.1932	0.0539
Residual	12.7312	5	2.5462		
Total	52.4009	8			

Coefficients

Attribute	Coef.	std	t(5)	p-value
Constant	2.397548	0.531899	4.507527	0.006355
[x1]	1.408402	0.564165	2.496438	0.054728
[x2]	1.642404	0.564165	2.911214	0.033353
<[x1][x2]>	-0.745165	0.797848	-0.933969	0.393193

Residuals analysis

Att. name: Err_Pred_lmreg_1

Full statistics

Statistics	
Average	0.0000
Median	-0.3363
Std dev. [Coef of variation]	1.2615 [-99999.0000]
MAD [MAD/STDDEV]	1.0433 [0.8270]
Min * Max [Full range]	-1.76 * 2.05 [3.81]
1st * 3rd quartile [Range]	-0.98 * 0.59 [1.57]
Skewness (std-dev)	0.3483 (0.7171)
Kurtosis (std-dev)	-0.8933 (1.3997)

Histogram

Values	Count	Percent	Histogr
x_<_-1.3759	1	11.11%	
-1.3759_=<_x_<_-0.9951	1	11.11%	
-0.9951_=<_x_<_-0.6144	1	11.11%	
-0.6144_=<_x_<_-0.2337	2	22.22%	
-0.2337_=<_x_<_0.1471	0	0.00%	
0.1471_=<_x_<_0.5278	0	0.00%	
0.5278_=<_x_<_0.9085	2	22.22%	
0.9085_=<_x_<_1.2893	0	0.00%	
1.2893_=<_x_<_1.6700	1	11.11%	
x>=_1.6700	1	11.11%	