APPLICATION OF PAIRS TRADING MODEL
TO EXCHANGE TRADED COFFEE FUTURES

by

Pratik Praful Bhandari
Bachelor of Technology, VJTI (University of Mumbai), 2013

and

Gursimran Manpreet Singh
Bachelor of Engineering, University of Mumbai, 2014

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN FINANCE

In the Master of Science in Finance Program
of the
Faculty
of
Business Administration

© Pratik Bhandari and Gursimran Singh 2016
SIMON FRASER UNIVERSITY
Fall 2016

All rights reserved. However, in accordance with the Copyright Act of Canada,
this work may be reproduced, without authorization, under the conditions for Fair
Dealing. Therefore, limited reproduction of this work for the purposes of private
study, research, criticism, review and news reporting is likely to be in accordance
with the law, particularly if cited appropriately.
APPROVAL

Name: Pratik Bhandari and Gursimran Singh

Degree: Master of Science in Finance

Title of Project: Application of Pairs Trading Model to Exchange Traded Coffee Futures

Supervisory Committee:

__________________________________________
Dr. Christina Atanasova
Senior Supervisor

__________________________________________
Dr. Evan Gatev
Second Reader

Date Approved: _____________________________
ABSTRACT

In this paper, we use the trading model as described by Kanamura, Rachev, & Fabozzi (2010) based on the pairs trading strategies using the Stochastic Spread Method and apply that to the exchange traded coffee futures (Generic ‘KC’ commodity). We also explain the first-time hitting density, (Linetsky 2004) for mean reverting process and apply this mathematical model to our data in order to find the results. In our empirical evidence, we test the real-time data obtained from Bloomberg in an Excel model based on co-integration approach to spread trading. We also show that the profits are consistent using the theoretical and empirical models and that profits depend on the mean reversion and volatility of the spread during the period under consideration.

Keywords: Pairs Trading; Coffee Futures; Spread Trading; Stochastic Spread Method; Mean Reversion.
DEDICATION

To my Family for their sacrifices and encouragement so I could pursue learning. To my friends who always kept me grounded & to Leo, for the therapeutic moist licks that kept me going....

- Gursimran Singh

I would like to dedicate this work to my Family, Uncle, Teachers, Industry Mentors and Friends for all their support.

- Pratik Praful Bhandari
ACKNOWLEDGEMENTS

We would like to thank Dr. Christina Atanasova and Dr. Evan Gatev for providing valuable insights, time, constructive comments and challenging us to consider ideas from all vantage points.

Last but not the least; we would like to thank our classmates with whom we have spent wonderful hours working together.
# TABLE OF CONTENTS

**APPROVAL** ................................................................................................................................. ii

**ABSTRACT** ................................................................................................................................. iii

**DEDICATION** ............................................................................................................................... iv

**ACKNOWLEDGEMENTS** ............................................................................................................... v

**TABLE OF CONTENTS** ............................................................................................................... vi

**LIST OF FIGURES** ....................................................................................................................... vii

**LIST OF TABLES** ......................................................................................................................... viii

1 - **INTRODUCTION** .................................................................................................................. 1

2 - **METHODOLOGY AND DATA** ............................................................................................. 6

2.1 Methodology ............................................................................................................................... 6

2.2 Data Description and Application of the Profit Model to Coffee Futures ............................... 8

2.3 Linetsky’s First Time Hitting Model ....................................................................................... 11

3 - **RESULTS** ............................................................................................................................. 12

4 - **EMPIRICAL EVIDENCE** ....................................................................................................... 13

5 - **CONCLUSION** ....................................................................................................................... 17

**APPENDICES** ............................................................................................................................. 18

Appendix 1: Spread Summary Characteristics ............................................................................. 18

Appendix 2: Regression Analysis between KC1 and KC2 .............................................................. 18

Appendix 3: MATLAB Code ........................................................................................................... 19

  3.1 Calculating Profits .................................................................................................................... 19

  3.2 AR(1) Model ........................................................................................................................... 20

**BIBLIOGRAPHY** ......................................................................................................................... 21
<table>
<thead>
<tr>
<th>List of Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1: Last Prices of KC1 and KC2 Commodity Coffee Futures</td>
</tr>
<tr>
<td>Figures 2 and 3: Expected Profits vs Std. Dev and Degree of Mean Reversion</td>
</tr>
<tr>
<td>Figure 4: Inputs to the Empirical Model</td>
</tr>
<tr>
<td>Figure 5: Outputs to the Empirical Model</td>
</tr>
<tr>
<td>Figure 6: Inputs to the Empirical Model – Strategy Inputs</td>
</tr>
<tr>
<td>Figure 7: Outputs to the Empirical Model – Strategy Performance</td>
</tr>
<tr>
<td>Figure 8: Summary of the entry and exit trades to be made as part of the</td>
</tr>
<tr>
<td>trading strategy</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Data Characteristics 8
Table 2: AR(1) Models for Coffee Bean Futures Price Spread 9
Table 3: Phillips-Perron Test for Futures Price Spreads 10
1 - INTRODUCTION

In this research project, we examine a very popular quantitative trading strategy called, ‘spread trading’ also known as ‘pairs trading’. It is one of the oldest and most widely used technique. The very concept behind this strategy is noticeably simple; that one can generate profits by identifying statistical mispricing of an asset based on their expected values or perhaps a convergence towards an expected mean value. One can say it gave birth to an interesting research avenue under quantitative trading called Statistical Arbitrage.

The usage of the word ‘arbitrage’ can be a bit deceptive because arbitrage trades are associated with being risk-free trades. However, spread trading is by no means risk free. There is a fair probability that the pairs in contention might not converge during the trading period and continue to diverge during the trading period or because of factors beyond the trader’s control. The risk would also be proportionate to the amount of capital the trader deploys for the particular trade. Since profits in the age of algorithms and electronic trading are squeezed, traders do tend to maximize returns by doing deploying large amount of capital.

There are four main methods to implement the pairs trading strategies as covered in the academic literature:

1. Distance Trading Method,
2. Stochastic Spread Method,
3. Combine Forecast Method and

The Distance Method

It is the most common method and the most referenced in the pairs trading strategies. The study by Gatev, Goetzmann and Rouwenhorst (2006) represents the pre-eminent contribution in the academic literature, which demonstrates
empirical evidence of excess returns from pairs trading when applied to US stocks. Their first research showed that the average annualized excess return is 12% of top pairs, and concluded that the pairs payoff is not strictly linked to a classical mean reversion effect. Abnormal returns are compensation to arbitrageurs for enforcing the “Law of One Price”. Do and Faff (2009) replicated the Gatev et al, (2006) methodology with the recent data. They found that the strategy is still profitable but it is declining. Bianchi, Drew and Zhu (2009) employed the pairs trading strategies on daily commodity futures returns and revealed that pairs trading in similarly related commodity futures earns statistically significant excess returns with commensurate volatility (Bianchi, Drew and Zhu 2009).

**Stochastic Spread Method**

A stochastic approach has been used in the pairs trading by Elliott (2005), Do et al, (2006), Kanamura, Rachev and Fabozzi (2010). Elliot (2005) proposed a mean reverting that is Gaussian Markov chain model for the spread. The appropriate investment decision is based on the predictions of the spread and is calibrated from market observations. Do et al, (2006) suggested that the long-term mean of the level differences in two stocks should not be constant and they proposed the stochastic residual spread method to pairs trading. Kanamura et al, (2010) applied the pairs trading strategy to energy futures market from 2000 to 2008 by using a mean reverting process of the futures price spread. They found that the stable profit can be made with the pairs trading but the profit of cross commodities may not be improved.

**Combine Forecast Method**

The method of combine forecast is described by Huck (2009, 2010). He used multi-criteria decision-making methods (MCDM) and neural networks methods to test pairs trading strategy by using S&P 100 stocks. These two methods are based on three stages: 1- forecasting, 2- ranking and 3- trading. The combine forecast method is developed without the reference to any equilibrium
model. Huck (2009, 2010) proposed that the method offers much more trading possibilities and could detect the “birth” of the divergence that other methods cannot achieve.

**Cointegration Method**

The Cointegration approach described by Lin, McRae and Gulati (2006), Schmidt (2008), and Puspaningrum (2009). Lin et al, (2006) developed a procedure that implants a minimum profit condition in the pairs trading strategy. Schmidt (2008) used the Johansen test for cointegration to identify pairs of stock and then mean-reverting residual spread modeled as a Vector-Error-Correction-Model (VECM). Puspaningrum (2009) tried to find the optimal pre-set boundaries for pairs trading strategy by using the cointegration method. The objective was to develop a quantitative method to assess the average trade duration, the average inter-trade interval and the average number of trades and then at the end of these assessments, the objective is to use them to find the optimal pre-set boundaries. In the term of maximizing the minimum total profit for co-integration error following an AR(1) process, the optimality is improved by assembling the cointegration technique, the cointegration coefficient weighted rule, and the mean first-passage time using an integral equation approach.

To trade using pairs trading strategy the first step is identifying a pair of instruments, which demonstrate a co-movement in prices. The basis for this can be, for example: two stocks, which are highly dependent on oil prices, would tend to have a mean reverting spread. The next step would be to have a formation period defined by (Gatev, Goetzmann and Rouwenhorst 2006) in their 2006 paper where they used a 12-month observation period or formation period. The trading period needs to be long enough to have opportunities to open and close trades and test the strategy but it cannot be too long because it is possible that the co-integration relationship between the two tested commodities will change. This is the time a trader would observe the movement of the two instruments and establish a certain relationship in their movement. Figure 1 shows how the two different
maturity futures with the same underlying depict co-movement in prices. Once the pair is established and the movement is observed we can calculate the spread. If the spread shows a significant divergence from the mean, a profitable trade can be locked with the appropriate positions.

**Figure 1: Last Prices of KC1 and KC2 Commodity Coffee Futures**

![Graph showing last prices of KC1 and KC2 Commodity Coffee Futures](image)

Until recently, the wide application of this strategy has been limited to equities. Kanamura, Rachev, & Fabozzi (2010) developed a profit model based on pairs trading strategy with an application to energy futures.

With this study, we intend to understand the theoretical and quantitative modelling of the pairs trading strategies using the profit model proposed by and subsequently test the model on agricultural futures. The profit model suggested in this research (Kanamura, Rachev, & Fabozzi, 2010) shows that one can obtain profits theoretically and they then verify it empirically as well. The paper tests the model on heating oil and natural gas futures (energy futures) and further suggests that it can be extended to any future. Our focus has been to apply this model on Coffee Bean futures (Generic ‘KC’ commodity’) because they are most liquid and most traded volatile futures on the ICE Exchange.
The model postulates that ultimately the profits depend on the mean reversion and volatility of the spread during the period under consideration. It models the spread as a stochastic process which is mean-reverting and also its first hitting time density.

**Section II** of this thesis explains the model in detail and explains the required mathematical background, first order mean reversion model and the rationale behind the trading strategy. We also explain the first-time hitting density, (Linetsky 2004) for mean reverting process. We apply this mathematical model to our data in order to find the results. There is a brief description about the data points we have used to calculate the spread and subsequently applied to the trading strategy to obtain profits.

In **Section III**, we consider a base case to test our model and the trading strategy. We display the results, which verify our assumptions that this particular model can be extended to commodity futures, the commodity in question being coffee beans. We also conduct a sensitivity analysis, where we observe the profits by testing it against an increase in degree of mean reversion and standard deviation.

**Section IV** depicts testing of real-time data obtained from Bloomberg in an Excel Model based on co-integration approach to spread trading. We find that trading coffee futures with different maturities can produce profits.

Finally, in **Section V**, we sum up and conclude our findings. In addition to that, we discuss briefly about its further scope and applications.
2 - METHODOLOGY AND DATA

2.1 Methodology

As per the Profit Model explained in the paper, if the two prices converge during the trading period, the position is closed at the time of the convergence. Otherwise, the position is forced to close at the end of the trading period. For example, suppose we denote by $P_{1,t}$ and $P_{2,t}$ the relatively high and low financial instrument prices respectively at time $t$ during the trading period. When the convergence of the price spread occurs at time $\tau$ during the trading period, the profit is the price spread at time 0 denoted by $x = P_{1,0} - P_{2,0}$. In contrast, when the price spread does not converge until the end of the trading period at time $T$, the profit stems from the difference between the spreads at times 0 and $T$ as $x - y = (P_{1,0} - P_{2,0}) - (P_{1,T} - P_{2,T})$, where $y$ represents the price spread at time $T$. Thus, spread trading produces a profit or a loss from the relative price movement, not the absolute price movement.

As presented, a profit model for this basic and most common spread trading strategy where the trade starts at time 0 and ends at time $T$, i.e., the trade is conducted once in the model. To do so, we need to take into account in the model:

1) The price spread movement and
2) The frequency of the convergence.

Here we use these basic modeling components to derive our profit model in order to formulate a general profit model for spread trading. In order to calculate the expected profits from the strategy we would require the probability density of converging $g(\gamma; x, T)$ and non-converging, $k(\gamma; x, T)$ scenarios respectively and also the first time hitting time density for the spread process.

The profit model, denoted by $r_p$, is thus derived from the following spread strategy explained:

Consider $x$ to be the price spread of two assets at time 0 when trading begins after taking a long and short position. First, let's consider the spread
convergence case in which price spreads at times 0 and \( \tau \) are fixed as \( x \) and 0, respectively. Assuming that the first hitting time density for the price spread process is \( f_{\tau \to 0}(t) \), the expected return \( r_{p,c} \) for this spread convergence is represented as the expectation value:

\[
r_{p,c} = x \int_0^\tau f_{\tau \to 0}(t) \, dt
\]

Next consider the case where there is a failure to converge during the trading period. This case holds if the price spread does not converge on 0 during the trading period and then reaches any price spread \( y \) at the end of the trading period. Note that \( y \) is greater than or equal to 0, otherwise, the process converges on 0 before the end of the trading period. This case is represented by the difference between two events:

i. When the process with initial value of \( x \) at time 0 reaches \( y \) at time \( T \) and

ii. When it arrives at \( y \) at time \( T \) after the process with initial value of \( x \) at time 0 converges on 0 at any time \( t \) during the trading period.

We denote the corresponding distribution functions by \( g(\gamma; x, T) \) and \( k(\gamma; x, T) \), respectively. In addition, the payoff of the trading strategy is given by \( x - y \).

A profit model for spread trading due to the failure to converge is expressed by the expectation value:

\[
r_{p,nc} = \int_0^\infty (x - y) \{ g(\gamma; x, T) - k(\gamma; x, T) \} \, d\gamma
\]

The probability density for the latter event, \( k(\gamma; x, T) \), is represented by the product of the first hitting time density \( f_{\tau \to 0}(t) \) and the density \( g(\gamma; x, T - t) \), meaning that the process reaches \( y' \) after the first touch because both events occur independently. Since \( t \) can be taken as any value during the trading period, the density function \( k(\gamma; x, T) \) is calculated as the integral of the product with respect to \( t \), as given by:

\[
k(\gamma; x, T) = \int_0^T f_{\tau \to 0}(t) g(\gamma; 0, T - t) \, dt
\]

A profit from spread trading for the failure to converge case is:
\[ r_{p,nc} = \int_0^\infty (x - \gamma) \left\{ g(\gamma; x, T) - \int_0^T f_{T-t_0}(t) g(\gamma; 0, T - t) \, dt \right\} \, dy \]

Thus, the total profit model for spread trading is expressed by:

\[ r_p = r_{p,c} + r_{p,nc} \]

### 2.2 Data Description and Application of the Profit Model to Coffee Futures

We use the daily closing prices of Coffee Bean (KC) commodity futures traded on the ICE Exchange. Each futures product includes six delivery months - from one month to six months. The time period covered is from October 24\textsuperscript{th} 2011 to November 10\textsuperscript{th} 2016. The data was obtained from Bloomberg and has following data characteristics as described in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>KC1 Comdty</th>
<th>KC2 Comdty</th>
<th>KC3 Comdty</th>
<th>KC4 Comdty</th>
<th>KC5 Comdty</th>
<th>KC6 Comdty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>149.16</td>
<td>151.77</td>
<td>154.48</td>
<td>157.02</td>
<td>159.40</td>
<td>161.59</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>143.80</td>
<td>146.88</td>
<td>149.80</td>
<td>152.40</td>
<td>155.18</td>
<td>157.13</td>
</tr>
<tr>
<td><strong>Max.</strong></td>
<td>250.80</td>
<td>253.60</td>
<td>254.80</td>
<td>255.10</td>
<td>254.30</td>
<td>253.80</td>
</tr>
<tr>
<td><strong>Min.</strong></td>
<td>101.50</td>
<td>104.60</td>
<td>106.75</td>
<td>109.05</td>
<td>111.35</td>
<td>114.45</td>
</tr>
<tr>
<td><strong>Std Dev.</strong></td>
<td>32.47</td>
<td>32.60</td>
<td>32.64</td>
<td>32.60</td>
<td>32.47</td>
<td>32.16</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.67</td>
<td>0.68</td>
<td>0.66</td>
<td>0.63</td>
<td>0.60</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>2.61</td>
<td>2.63</td>
<td>2.59</td>
<td>2.55</td>
<td>2.50</td>
<td>2.48</td>
</tr>
</tbody>
</table>

Futures contracts have a similar term structure as interest rates. It is either upward sloping also known as contango (futures price is above the current spot price), or downward sloping also known as backwardation (futures price is below the current spot price). Thus, coffee futures prices that have different maturities converge on the flat term structure, even though price deviates from the others due to the backwardation or contango. These attributes that we observe for the term structure of coffee futures prices may be useful in obtaining the potential profit in coffee futures markets using the convergence of the price spread between the
two correlated assets as in spread trading. More importantly, the transition of price
spreads from backwardation or contango to the flat term structure can be
considered as a mean reversion of the spread.

In order to support this conjecture, we estimate the following autoregressive
1 lag AR(1) model for price spreads \((P^i_t - P^j_t)\) for \(i\) and \(j\) month coffee futures \((i < j)\).

\[
P^i_t - P^j_t = C_{ij} + \rho_{ij} (P^i_{t-1} - P^j_{t-1}) + \epsilon_t
\]

The results are reported as follows in the following Table 2:

<table>
<thead>
<tr>
<th>Table 2: AR(1) Models for Coffee Bean Futures Price Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Coefficient</td>
</tr>
<tr>
<td>Std. Error</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Coefficient</td>
</tr>
<tr>
<td>Std. Error</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Coefficient</td>
</tr>
<tr>
<td>Std. Error</td>
</tr>
</tbody>
</table>

The important step of the strategy consists of identifying potential pairs.
First, to test for a unit root in the individual futures price series, the Phillips-Perron
tests are applied. All the unit root tests are performed, for all futures prices and
spreads in their levels; and the tests are performed with their first difference values
if found not stationary for the actual prices.

PP tests for the futures prices, the null hypothesis is not rejected at the 1%
significance, indicating that all variables are not stationary. Using the PP test, the
null hypothesis is rejected for all first difference equations at the 1% level of
significance. As shown in table 3, PP tests for the futures spreads, the null
hypothesis is rejected at the 1% significance, indicating that all variables are
stationary and that the spreads are mean-reverting.
Table 3: Phillips-Perron Test for Futures Price Spreads

<table>
<thead>
<tr>
<th>Two maturity months</th>
<th>Phillip-Perron test statistic</th>
<th>Test critical values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>KC12</td>
<td>-6.572</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC13</td>
<td>-6.146</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC14</td>
<td>-6.057</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC15</td>
<td>-5.933</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC16</td>
<td>-5.479</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC23</td>
<td>-4.586</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC24</td>
<td>-4.536</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC25</td>
<td>-4.942</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC26</td>
<td>-5.016</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC34</td>
<td>-4.620</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC35</td>
<td>-4.856</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC36</td>
<td>-4.827</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC45</td>
<td>-4.972</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC46</td>
<td>-4.615</td>
<td>-3.965</td>
</tr>
<tr>
<td>KC56</td>
<td>-4.303</td>
<td>-3.965</td>
</tr>
</tbody>
</table>

As can be seen from the results of the AR(1) model, AR(1) coefficients for all combinations of the price spreads are statistically significant and greater than 0 and less than 1. In addition, the Phillips-Perron tests for the levels of price spreads all reject the existence of unit roots because the Phillips-Perron test statistics are less than three test critical values.

Moreover, based on the model theorised in the paper and the characteristics of the price spreads for the coffee futures that we observe, it can be concluded that a mean-reverting model with a long term mean can be applicable to coffee futures price spreads ($S_t$) as described below:

\[ dS_t = \kappa(\theta - S_t)dt + \sigma dW_t \]

Where, $\kappa = -\ln(\rho_{ij})$; $\theta = \frac{c_{ij}}{1-\rho_{ij}}$; $\sigma = \sigma_{ij} \sqrt{\frac{-2\ln(\rho_{ij})}{1-\rho_{ij}^2}}$
2.3 Linetsky’s First Time Hitting Model

Linetsky provided an analytical model for first time hitting probability density for mean reverting processes which is applicable in modelling interest rates, stochastic volatility, credit spreads and convenience yields. The first time hitting density for a process moving from $x$ to $0$ is given by:

$$f_{\tau_x \to 0}(t) = \sum_{n=1}^{\infty} C_n \lambda_n e^{-\lambda_n t}$$

Where $\lambda_n$ and $C_n$ have large $n$-asymptotics given by:

$$\lambda_n = \kappa \left(2k_n - \frac{1}{2}\right)$$

$$k_n = n - \frac{1}{4} + \frac{y\sqrt{2}}{\pi} \left[-\frac{1}{4} + \frac{y^2}{2\pi^2}\right]$$

$$C_n = [-1^{(n+1)}2\sqrt{k_n}/(2k_n - 0.5)(\pi\sqrt{k_n} - 2\frac{1}{2}y)] \ast e^{0.25(\bar{x}^2 - \bar{y}^2)} \ast \cos(\bar{x}\sqrt{2k_n} - \pi k_n + \pi/4)$$
3 - RESULTS

We test the model for the equations for profit derived in Section II. To do that we consider the spread between the 2-month, KC2, and 1-month, KC1 maturity coffee bean futures. We observe that the initial spread is 2.8 at the beginning of our trading period and it hits 0 for the first time at the 335th observation. Hence, we are testing for the period where our spread converges to zero and the trading period effectively ends there as we unwind our position.

We consider our long term mean, $\theta = 0$, calculate kappa, $\kappa$, as $-\ln(\rho)$ which gives us a value of 0.0692 and take standard deviation as estimated from the AR(1) model, $\sigma = 0.249$. In addition to that we assume that we begin trading at a time when the spreads are twice the standard deviation therefore we get $x = 0.5141$. Using the equation for profit an as defined in the earlier section we get a profit of 5.9858. This reaffirms the fact that the model does apply both theoretically and empirically to coffee futures as the spreads show both mean reversion and price volatility.

The equations for profit are expressed as functions of $\kappa$ and $\sigma$ which means that the profits are influenced highly by degree of mean reversion and volatility. Therefore, higher the $\sigma$ and $\kappa$ higher should be profits. Looking at the sensitivity analysis of these two characteristics separately we find that increase in both, in isolation, increases the profits. These are exhibited by the figures below:

Figures 2 and 3: Expected Profits vs Std. Dev and Degree of Mean Reversion
4 - EMPIRICAL EVIDENCE

It is important for any theoretical model to be tested with the current data in order to check if the trading strategy is still significantly profitable. We have used a spreadsheet model implemented on the co-integration method as described in the paper with an extension to the coffee futures spreads (data collected from the Bloomberg) for the sample period starting from October 2011 to November 2016.

Firstly, we input a pair, date ranges and review its characteristics to find a good mean reversion candidate. We look for a good regression fit and high number of zero crossings in the residual spread series.

Figure 4: Inputs to the Empirical Model

Input:

<table>
<thead>
<tr>
<th>Pair:</th>
<th>A (Indep.)</th>
<th>B (Dep.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KC1 Comdty</td>
<td>KC2 Comdty</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Period:</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/24/2011</td>
<td>11/10/2016</td>
<td></td>
</tr>
</tbody>
</table>

Generate Pair Characteristics

Figure 5: Outputs to the Empirical Model

Output (Characteristics of the Pair for Regression Period):

a. Regression Fit for natural log of prices: Chart 1

\[ y = 0.986x + 0.0871 \]

\[ R^2 = 0.9977 \]
b. Residual Spread Series:

No. of crossings around 0 (Higher is better): 74
Avg. crossing period (Smaller is better): 16

![Residual spreads chart](chart2)

-0.03 -0.02 -0.01 0 0.01 0.02 0.03 0.04 0.05 0.06 0.07

Residual spreads
-1 Std. Dev. -2 Std. Dev.
+1 Std. Dev. +2 Std. Dev.

c. Closing Prices:

![Closing Prices chart](chart3)

0 50 100 150 200 250 300

KC2 Comdty KC1 Comdty
Secondly, using the setup, we input the trading parameters and test the pair trading strategies as follows:

**Figure 6: Inputs to the Empirical Model – Strategy Inputs**

<table>
<thead>
<tr>
<th>Input:</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>When residual spread is positive, go:</td>
<td>Long</td>
<td>Short</td>
</tr>
<tr>
<td>When residual spread is negative, go:</td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>From</td>
<td>To</td>
<td></td>
</tr>
<tr>
<td>07/01/2015</td>
<td>28/06/2016</td>
<td></td>
</tr>
<tr>
<td>Std. Dev. multiplier for residual spread:</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Specify no. of shares or amount for A:</td>
<td>50,000</td>
<td></td>
</tr>
<tr>
<td>No. of shares for B:</td>
<td>Dollar Matching</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7: Outputs to the Empirical Model – Strategy Performance**

**Output (Strategy performance for Trading Period):**

<table>
<thead>
<tr>
<th>a. Summary:</th>
<th>View Trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Profit (Loss) if open positions are closed on last day: $</td>
<td>3,909.60</td>
</tr>
<tr>
<td>Profit (Loss) on closed positions: $</td>
<td>3,909.60</td>
</tr>
<tr>
<td>Is position open on last day:</td>
<td>No</td>
</tr>
<tr>
<td>Maximum negative excursion of any trade: $</td>
<td>(292.60)</td>
</tr>
<tr>
<td>Maximum positive excursion of any trade: $</td>
<td>812.10</td>
</tr>
<tr>
<td>No. of crossings around 0:</td>
<td>17</td>
</tr>
<tr>
<td>Avg. crossing period:</td>
<td>21</td>
</tr>
</tbody>
</table>
Figure 8: Summary of the entry and exit trades to be made as part of the trading strategy

<table>
<thead>
<tr>
<th>Date</th>
<th>No. of Trading Days</th>
<th>Price (A)</th>
<th>Price (B)</th>
<th>Qty (A)</th>
<th>Qty (B)</th>
<th>Entry/Exit</th>
<th>Cumulative Cashflow</th>
<th>Individual P&amp;L</th>
<th>Cumulative P&amp;L</th>
</tr>
</thead>
<tbody>
<tr>
<td>28/04/2015</td>
<td>Entry</td>
<td>138.15</td>
<td>138.55</td>
<td>-362</td>
<td>361</td>
<td>(6.25)</td>
<td>3,659.75</td>
<td>836.50</td>
<td>3,909.60</td>
</tr>
<tr>
<td>20/05/2015</td>
<td>Exit</td>
<td>17</td>
<td>136</td>
<td>138.7</td>
<td>362</td>
<td>-361</td>
<td>838.70</td>
<td>832.45</td>
<td>832.45</td>
</tr>
<tr>
<td>20/06/2015</td>
<td>Entry</td>
<td>127.65</td>
<td>122.45</td>
<td>-392</td>
<td>378</td>
<td>-378</td>
<td>809.20</td>
<td>832.45</td>
<td>1,668.95</td>
</tr>
<tr>
<td>19/11/2015</td>
<td>Exit</td>
<td>65</td>
<td>119.9</td>
<td>122.2</td>
<td>378</td>
<td>-417</td>
<td>859.75</td>
<td>1,668.95</td>
<td>1,668.95</td>
</tr>
<tr>
<td>15/12/2015</td>
<td>Entry</td>
<td>115.65</td>
<td>119.9</td>
<td>432</td>
<td>417</td>
<td>37.50</td>
<td>1,706.45</td>
<td>1,329.45</td>
<td>2,998.40</td>
</tr>
<tr>
<td>16/12/2015</td>
<td>Exit</td>
<td>2</td>
<td>118.1</td>
<td>119.25</td>
<td>-432</td>
<td>417</td>
<td>1,291.95</td>
<td>2,998.40</td>
<td>1,329.45</td>
</tr>
<tr>
<td>17/12/2015</td>
<td>Entry</td>
<td>117.8</td>
<td>118.3</td>
<td>-324</td>
<td>422</td>
<td>24.60</td>
<td>3,023.00</td>
<td>2,991.30</td>
<td>3,023.00</td>
</tr>
<tr>
<td>22/06/2016</td>
<td>Exit</td>
<td>129</td>
<td>136.95</td>
<td>139.7</td>
<td>424</td>
<td>-422</td>
<td>886.60</td>
<td>911.20</td>
<td>3,909.60</td>
</tr>
</tbody>
</table>
5 - CONCLUSION

In this thesis, we applied a general profit model on exchange traded coffee futures. We chose coffee futures contracts, as it is the most traded commodity in volume after oil. The model is based on spread trading strategy, which focusses on the stochastic movement of the spread. The profits for this can be calculated given that we know the initial spread and the first hitting time probability density.

We examine the spread between two different maturity futures contracts for mean reversion and volatility. We found the spreads show a significant degree of mean reversion and high volatility. We successfully implemented the model using a base case as described in Section III and conclude that the profits are enhanced by the degree of mean reversion and increasing volatility.

We tested the profitability using real time data and a co-integration model on Excel that reaffirmed the results obtained from the theoretical model that the profits were enhanced by degree of mean reversion and volatility.
APPENDICES

Appendix 1: Spread Summary Characteristics

Appendix 2: Regression Analysis between KC1 and KC2
Appendix 3: MATLAB Code

3.1 Calculating Profits

% Initial Values:
% These values are our values for which we test our model. We have taken
% the case of KC12. We assume that the long term mean is 0 for our base
% case and trading period of 335 days where it hits 0 for the first time.
% However, we have to be vary of taking extremely large values as the
% spread might not show mean reversion or trend for that large period of time.

T = 300;
C = 0.157;
rho = 0.939338;
sigma_ij = 0.249195;
kappa = -log(rho);
theta = 0
sigma = sigma_ij*sqrt((-2*log(rho))/(1-(rho^2)));
x = 2*sigma
xbar = (sqrt(2*kappa)*(x-theta))/(sigma);
ybar = (sqrt(2*kappa)*(-theta))/(sigma);

%%pre-allocating the matrix to obtain efficiency
k = zeros(1,T);
lambda = zeros(1, T);
c = zeros(1,T);
return_converge = zeros(1,T);

% we run iterations to solve the Linetsky first hitting time density function.
for i = 1:200
\[ k(i) = i - 0.25 + \frac{(y_{\bar{b}}^2)}{\pi^2} + (y_{\bar{b}} \cdot \sqrt{2}) \cdot (\sqrt{(i - 0.25 + \frac{(y_{\bar{b}}^2)}{2 \cdot \pi^2})}); \]

\[ \lambda(i) = \kappa \cdot (2k(i) - 0.5); \]

\[ c(i) = \left(\frac{(-1)^{(i+1)} \cdot 2 \cdot (\sqrt{k(i)})}{(2 \cdot k(i) - 0.5)^2} \cdot \left((\pi \cdot \sqrt{k(i)}) - (2^{-0.5} \cdot y_{\bar{b}})\right)\right) \cdot \exp(0.25 \cdot ((x_{\bar{b}}^2) - (y_{\bar{b}}^2))) \cdot \cos((x_{\bar{b}} \cdot \sqrt{2k(i)}) - (\pi \cdot k(i)) + (0.25 \cdot \pi)); \]

\[ \text{return\_converge}(i) = c(i) \cdot (1 - \exp(-\lambda(i) \cdot T)); \]

\[ \text{end} \]

% Thus by convergence the profits are calculated as:

\[ \text{profits\_converge} = \sum(\text{return\_converge}) \cdot x \]

### 3.2 AR(1) Model

%AR model estimation: Using the example for KC12

\[ y1 = \text{KC12}; \]

\[ \% \text{MDL} = \text{arima('ARLags', 1)}; \]

\[ \text{EstMdl} = \text{estimate(MDL, y1)}; \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.157155</td>
<td>0.0240787</td>
<td>6.5267</td>
</tr>
<tr>
<td>AR{1}</td>
<td>0.939338</td>
<td>0.00502722</td>
<td>186.85</td>
</tr>
<tr>
<td>Variance</td>
<td>0.249195</td>
<td>0.00292621</td>
<td>85.1594</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY