

MANAGING AN AGRICULTURAL  
COMMODITIES PORTFOLIO IN SOUTH AFRICA  
WITH PAIRS TRADING

ANDRÉ HEYMANS

THESIS SUBMITTED TO  
THE CENTRE FOR BUSINESS MATHEMATICS AND INFORMATICS OF  
THE NORTH-WEST UNIVERSITY (POTCHEFSTROOM CAMPUS)  
IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF  
PHILOSOPHIAE DOCTOR (RISK MANAGEMENT)

**Supervisor:**                      **Professor Doctor Paul Styger**

**Potchefstroom**  
**2008**

# ACKNOWLEDGEMENTS

I thank God Almighty in Whose grace I live every day

To Sabrina, whose love and support have made it possible to drag through the long nights in the seemingly never-ending process of finishing this thesis.

To my supervisor, Professor Paul Styger, for his leadership, friendship and patience, in times that I forgot what my own research was about.

Aan my ouers, vir hul liefde, ondersteuning en eindelose belangstelling ten spyte van die herhaalde woorde “ek weet nie hoe lank ek nog aan my proefskrif gaan werk nie”.

Aan Professor Francois van Graan vir sy hulp en verduideliking van ARIMA modelle aan my.

Aan Professor Hennie Venter vir sy hulp en verduideliking van GARCH en UCM.

Aan Gert, vir sy vriendskap en gedurige ondersteuning.

Aan Lukas, Adriaan en Werner, sonder wie se kennis van die Suid Afrikaanse landbou kommoditeite mark ek nie hierdie proefskrif sou kon voltooi nie.

Aan Professor Christo Auret wie aan my die data gegee het vir die voltooiing van hierdie proefskrif.

To Professor Riaan de Jongh at the Centre for Business Mathematics and Informatics for the financial help through my National Research Foundation bursary.

André Heymans

2008

## ABSTRACT

Although pairs trading is well known among South African agricultural commodity traders, there are no comprehensive documented accounts for the selection and trading of agricultural commodity pairs in South Africa. The majority of agricultural commodity pairs traders take positions based on their personal view of price movements, without testing for a statistical relationship between the paired commodities that will guarantee that their prices will move back to a common mean.

To remedy this lack of method regarding the pairs selection and pairs trading processes, a comprehensive pairs selection process was developed and is documented in this thesis. During the pairs selection process, several agricultural commodities were put through a rigorous evaluation process to test for any long-run statistical relationships between them. This was done to ensure that only pairs with stable long-run statistical relationships were included in the final pairs portfolio that was compiled.

In order to test the profitability of this pairs portfolio, several fundamental and technical indicators were used to determine entry and exit points. Although some of these indicators did not render satisfactory results, the RSI and Bollinger bands succeeded in realising an acceptable profit.

*Keywords:* Fundamental analysis, GARCH, agricultural commodities trader, co-integration, pairs trading, pairs selection process, seasonal patterns, technical analysis, UCM, X12.

## OPSOMMING

Hoewel 'pairs trading' nie 'n nuwe begrip is onder Suid-Afrikaanse landboukommoditeitshandelaars nie, bestaan daar nie 'n volledig gedokumenteerde proses vir die seleksie en verhandeling van landbou kommoditeite in Suid Afrika nie. Die meerderheid landboukommoditeitshandelaars neem posisies wat berus op 'n persoonlike mening oor die rigting en die grootte van prysbewegings, sonder om seker te maak of die twee kommoditeit reekse wat hy as paar gebruik, tog 'n statistiese verbandskap tussen hulle het wat sal verseker dat hierdie twee se pryse terug beweeg na 'n gemeenskaplike gemiddeld.

Ten einde 'n oplossing te vind vir hierdie tekort aan struktuur aangaande die pare seleksie proses, word 'n volledige strukturele pare seleksie proses ontwikkel, en voorgehou in hierdie proefskrif. Tydens die pare seleksie proses word verskillende landbou kommoditeite deeglik deur 'n streng keuringsproses geneem om uiteindelik langtermyn pare met stabiele statistiese verhoudings in 'n finale portefeulje saam te vat.

Om die winsgewendheid van hierdie pare portefeulje te toets word daar gebruik gemaak van verskeie fundamentele en tegniese indikatore om intree en uittree geleenthede te bepaal. Hoewel daar met die gebruik van sommige van die metodes nie daarin geslaag kon word om 'n noemenswaardige wins te toon nie, kon die 'RSI' en 'Bollinger bands' wel gebruik word om 'n aansienlike wins te behaal.

*Sleutelbegrippe:* Fundamentele analise, GARCH, graan handelaar, ko-integrasie, 'Pairs trading', pare seleksie proses, seisoenale patrone, tegniese analise, UCM, X12.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
OPSOMMING	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	x
LIST OF GRAPHS	x
LIST OF TABLES	xiii
CHAPTER 1: PAIRS TRADING: AN INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 PAIRS TRADING: AN OVERVIEW	2
1.2.1 Introduction	2
1.2.2 A Brief Overview of the History of Pairs Trading	2
1.2.3 What is Pairs Trading?	3
1.3 PROBLEM STATEMENT	5
1.4 RESEARCH AIMS AND OBJECTIVES	5
1.5 METHODOLOGY	6
1.6 CHAPTER OUTLINE	6
1.7 NOTES TO THE READER	8
1.7.1 Publications	8
1.7.2 Miscellaneous	9
CHAPTER 2: AGRICULTURAL COMMODITIES TRADING FUNDAMENTAL ANALYSIS	10
2.1 INTRODUCTION	10
2.2 THE HISTORY OF THE COMMODITIES TRADED ON SAFEX	11
2.2.1 Introduction	11
2.2.2 Maize	11
2.2.3 Wheat	13
2.2.4 Sunflower	14
2.2.5 Soybeans	16
2.2.6 Conclusion	17
	iv

2.3 AGRICULTURAL COMMODITY FUTURES PRICES	17
2.3.1 Introduction	17
2.3.2 The Supply and Demand of Commodities	18
2.3.3 Other Factors Influencing Derivative Contract Pricing	21
2.3.3.1 Introduction	21
2.3.3.2 The Basis	22
2.3.3.3 Contango and Backwardation Markets	22
2.3.3.4 Factors Influencing the Pricing of Futures Contracts	23
2.3.4 Conclusion	24
2.4 FUTURES EXCHANGES	25
2.4.1 Introduction	25
2.4.2 CBOT	25
2.4.3 Chicago Mercantile Exchange (CME)	26
2.4.4 New York Board of Trade (NYBOT)	26
2.4.5 Euronext.liffe	26
2.4.6 Tokyo Grain Exchange (TGE)	27
2.4.7 Sydney Futures Exchange (SFE)	27
2.4.8 South African Futures Exchange (SAFEX)	28
2.4.9 Conclusion	28
2.5 SUMMARY	29
CHAPTER 3: TECHNICAL ANALYSIS	30
3.1 INTRODUCTION	30
3.2 MARKET STRENGTH INDICATORS	30
3.2.1 Introduction	30
3.2.2 Overbought and Oversold	31
3.2.3 Momentum	31
3.2.4 The RSI	32
3.2.5 The Chande Momentum Oscillator	34
3.2.6 The Stochastic Oscillator	35
3.2.7 Conclusion	38
3.3 MOVING AVERAGE INDICATORS	38
3.3.1 Introduction	38
3.3.2 Volatility and Standard Deviation	39
3.3.2.1 Introduction	39
3.3.2.2 SMA	39
3.3.2.3 EWMA	40
3.3.2.4 Conclusion	41
3.3.3 Moving Averages as Indicator	42
3.3.4 Bollinger Bands	43
3.3.5 MACD	46
3.3.6 Conclusion	48
3.4 SUMMARY	48

CHAPTER 4: STATISTICAL MEASURES USED FOR TESTING SEASONALITY	50
4.1 INTRODUCTION	50
4.2 SEASONALITY	51
4.2.1 Introduction	51
4.2.2 Literature Review	51
4.2.3 Conclusion	53
4.3 AN INTRODUCTION TO ARMA AND ARIMA	53
4.4 AN INTRODUCTION TO GARCH	55
4.5 AN INTRODUCTION TO STA	61
4.6 SUMMARY	65
CHAPTER 5: TESTING FOR SEASONALITY	66
5.1 INTRODUCTION	66
5.2 DATA	67
5.2.1 Introduction	67
5.2.2 The Data	68
5.2.3 Conclusion	69
5.3 MODELLING FOR SEASONAL PATTERNS WITH GARCH	69
5.3.1 Introduction	69
5.3.2 Methodology	70
5.3.2.1 White Maize	72
5.3.2.1.1 Data Results	72
5.3.2.1.2 Model Selection	73
5.3.2.1.3 The Model	74
5.3.2.1.4 Model Results	76
5.3.2.1.5 Final Results – White Maize	78
5.3.2.2 Yellow Maize	80
5.3.2.2.1 Data Results	80
5.3.2.2.2 Model Selection	81
5.3.2.2.3 The Model	82
5.3.2.2.4 Model Results	83
5.3.2.2.5 Final Results – Yellow Maize	84
5.3.3 Conclusion	85
5.4 MODELLING FOR SEASONAL PATTERNS WITH X12	86
5.4.1 Introduction	86
5.4.2 Methodology	88
5.4.2.1 Introduction	88
5.4.2.2 The General Model	88
5.4.2.3 White Maize	92
5.4.2.3.1 Model Results	92

5.4.2.3.2 <i>Final Results – White Maize</i>	95
5.4.2.4 Yellow Maize	97
5.4.2.4.1 <i>Model Results</i>	97
5.4.2.4.2 <i>Final Results – Yellow Maize</i>	99
5.4.2.5 White Maize vs. Yellow Maize	100
5.4.2.6 Testing for Seasonality in the South African Maize Data with X12	103
5.4.3 Conclusion	106
5.5 MODELLING FOR SEASONAL PATTERNS WITH UCM	107
5.5.1 Introduction	107
5.5.2 Methodology	107
5.5.2.1 Estimating the Model – General Model Specification	107
5.5.2.1.1 <i>Modelling the Trend Component</i>	108
5.5.2.1.2 <i>Modelling the Cyclical Component</i>	108
5.5.2.1.3 <i>Modelling the Seasonal Component</i>	109
5.5.2.2 White Maize	110
5.5.2.2.1 <i>The Model</i>	110
5.5.2.2.2 <i>Model Results</i>	111
5.5.2.2.3 <i>Final Results – White Maize</i>	112
5.5.2.3 Yellow Maize	113
5.5.2.3.1 <i>The Model</i>	113
5.5.2.3.2 <i>Model Results</i>	115
5.5.2.3.3 <i>Final Results – Yellow Maize</i>	116
5.5.2.4 White Maize vs. Yellow Maize	116
5.5.2.5 Testing for Seasonality in the South African Maize Data with UCM	117
5.5.3 Conclusion	122
5.6 SUMMARY	123
CHAPTER 6: THE PAIRS SELECTION PROCESS	125
6.1 INTRODUCTION	125
6.2 PAIRS SELECTION: STATISTICAL MEASURES	126
6.2.1 Introduction	126
6.2.2 Statistical Measures	127
6.2.2.1 Co-integration	128
6.2.2.2 Utilising Co-integration in Selecting Long-run Pairs	130
6.2.2.2.1 <i>Introduction</i>	130
6.2.2.2.2 <i>Testing for Unit Roots: Methodology</i>	131
6.2.2.2.3 <i>Testing for Unit Roots: Results</i>	131
6.2.2.2.4 <i>Testing for Co-integration: Methodology</i>	134
6.2.2.2.5 <i>Testing for Co-integration: Results</i>	137
6.2.2.2.6 <i>Conclusion</i>	140
6.2.2.3 Pearson Correlation	140
6.2.2.4 Utilising Pearson’s Correlation to Validate Long-run Pairs	143
6.2.2.4.1 <i>Introduction</i>	143
6.2.2.4.2 <i>Testing for Pearson’s Correlation: Methodology</i>	143
6.2.2.4.3 <i>Testing for Pearson’s Correlation: Results</i>	145
6.2.2.4.4 <i>Conclusion</i>	146
6.2.2.5 Spearman Rank Correlation	147
6.2.2.6 Utilising Spearman Rank Correlation to Validate Long-run Pairs	148
6.2.2.6.1 <i>Introduction</i>	148

6.2.2.6.2 <i>Testing for Spearman's Rank Correlation: Methodology</i>	148
6.2.2.6.3 <i>Testing for Spearman's Rank Correlation: Results</i>	149
6.2.2.6.4 <i>Conclusion</i>	150
6.2.2.7 <i>Beta</i>	151
6.2.2.8 <i>Utilising Beta to Validate Long-run Pairs</i>	152
6.2.2.8.1 <i>Introduction</i>	152
6.2.2.8.2 <i>Testing for Beta: Methodology</i>	152
6.2.2.8.3 <i>Testing for Beta: Results</i>	153
6.2.2.8.4 <i>Conclusion</i>	155
6.2.3 <i>Conclusion</i>	156
<b>6.3 PAIRS SELECTION: DETERMINISTIC SEASONALITY MEASURES</b>	<b>156</b>
6.3.1 <i>Introduction</i>	156
6.3.2 <i>Testing for Pairs with X12</i>	157
6.3.2.1 <i>Introduction</i>	157
6.3.2.2 <i>The Henderson-curve as Pairs Selection Tool</i>	157
6.3.2.3 <i>Seasonal Factors as Pairs Selection Tool</i>	162
6.3.3 <i>Conclusion</i>	167
<b>6.4 PAIRS SELECTION: STOCHASTIC SEASONALITY MEASURES</b>	<b>167</b>
6.4.1 <i>Introduction</i>	167
6.4.2 <i>Testing for pairs with UCM: Methodology</i>	168
6.4.3 <i>Conclusion</i>	171
<b>6.5 SUMMARY</b>	<b>172</b>
<b>CHAPTER 7: MANAGING AN AGRICULTURAL COMMODITIES PORTFOLIO WITH PAIRS TRADING</b>	<b>175</b>
<b>7.1 INTRODUCTION</b>	<b>175</b>
<b>7.2 DATA</b>	<b>176</b>
<b>7.3 MANAGING A PAIRS PORTFOLIO</b>	<b>180</b>
7.3.1 <i>Introduction</i>	180
7.3.2 <i>Methodology</i>	182
7.3.2.1 <i>Introduction</i>	182
7.3.2.2 <i>Assumptions and Trading Rules</i>	182
7.3.2.2.1 <i>Introduction</i>	182
7.3.2.2.2 <i>Entry and Exit Points</i>	182
7.3.2.2.3 <i>Identifying Entry and Exit points</i>	184
7.3.2.2.4 <i>Brokerage and Margins</i>	185
7.3.2.3 <i>Steps Followed during the Test Trading Process</i>	186
<b>7.4 RESULTS</b>	<b>187</b>
7.4.1 <i>Introduction</i>	187
7.4.2 <i>Final Results of Approach One: All indicators</i>	187
7.4.3 <i>Final Results of Approach Two: Technical Analysis Tools</i>	189
7.4.3.1 <i>Utilising the RSI to Identify Entry and Exit Points</i>	189
7.4.3.2 <i>Utilising the Stochastic Oscillator to Identify Entry and Exit Points</i>	191
7.4.3.3 <i>Utilising Bollinger Bands to Identify Entry and Exit Points</i>	192
7.4.3.4 <i>Utilising MACD to Identify Entry and Exit Points</i>	194

7.4.3.5 Utilising MA to Identify Entry and Exit Points	195
7.4.4 Conclusion	196
7.5 SUMMARY	197
CHAPTER 8: CONCLUSION	200
8.1 INTRODUCTION	200
8.2 AIM OF THE RESEARCH PROJECT	200
8.3 STUDY REVIEW: FUNDAMENTAL ANALYSIS	200
8.4 STUDY REVIEW: TECHNICAL ANALYSIS	201
8.5 STUDY REVIEW: MEASURES USED TO TEST FOR SEASONALITY	201
8.6 STUDY REVIEW: TESTING FOR SEASONALITY	201
8.7 STUDY REVIEW: PAIRS SELECTION	202
8.8 STUDY REVIEW: TRADING AGRICULTURAL COMMODITY PAIRS	203
8.9 CONCLUSION	204
8.10 RECOMMENDATIONS FOR FURTHER STUDY	205
REFERENCES	207
APPENDICES	219
INDEX	229

## LIST OF FIGURES

### CHAPTER 3

FIGURE 3.1	RSI BULLISH AND BEARISH DIVERGENCES	33
FIGURE 3.2	IDENTIFYING CMO ENTRY AND EXIT POSITIONS	35
FIGURE 3.3	THE STOCHASTIC OSCILLATOR	37
FIGURE 3.4	MOVING AVERAGES	43
FIGURE 3.5	BOLLINGER BANDS	45
FIGURE 3.6	THE MACD INDICATOR	47

### CHAPTER 6

FIGURE 6.1	CORRELATION ANALYSIS	141
------------	----------------------	-----

## LIST OF GRAPHS

### CHAPTER 1

GRAPH 1.1	ENTRY AND EXIT POINTS IN A PAIRS TRADE	3
-----------	--	---

### CHAPTER 5

GRAPH 5.1	MONTHLY AVERAGE SPOT PRICE SERIES OF WHITE MAIZE VS. HENDERSON-CURVE	96
GRAPH 5.2	MONTHLY AVERAGE SPOT PRICE SERIES OF WHITE MAIZE VS. SEASONAL FACTORS SERIES	97
GRAPH 5.3	MONTHLY AVERAGE SPOT PRICE SERIES OF YELLOW MAIZE VS HENDERSON-CURVE	99
GRAPH 5.4	MONTHLY AVERAGE SPOT PRICE SERIES OF YELLOW MAIZE VS. SEASONAL FACTORS SERIES	100
GRAPH 5.5	WHITE MAIZE HENDERSON-CURVE VS. YELLOW MAIZE HENDERSON-CURVE	101
GRAPH 5.6	WHITE MAIZE SEASONAL FACTORS SERIES VS. YELLOW MAIZE SEASONAL FACTORS SERIES	102
GRAPH 5.7	MONTHLY AVERAGE SPOT PRICE SERIES OF WHITE MAIZE VS HIDDEN SEASONAL	113
GRAPH 5.8	MONTHLY AVERAGE SPOT PRICE SERIES OF YELLOW MAIZE VS HIDDEN SEASONAL	116
GRAPH 5.9	WHITE MAIZE HIDDEN SEASONAL VS. YELLOW MAIZE HIDDEN SEASONAL	117

## CHAPTER 6

GRAPH 6.1	WMSPOT – WM 1 AWAY	158
GRAPH 6.2	WMSPOT – WM 2 AWAY	158
GRAPH 6.3	WMSPOT – WM 3 AWAY	158
GRAPH 6.4	WMSPOT – YMSPOT	158
GRAPH 6.5	WMSPOT – YM 1 AWAY	159
GRAPH 6.6	WMSPOT – YM 2 AWAY	159
GRAPH 6.7	WMSPOT – YM 3 AWAY	159
GRAPH 6.8	WM 1 AWAY – WM 2 AWAY	159
GRAPH 6.9	WM 1 AWAY – WM 3 AWAY	159
GRAPH 6.10	WM 1 AWAY – YMSPOT	159
GRAPH 6.11	WM 1 AWAY – YM 1 AWAY	159
GRAPH 6.12	WM 1 AWAY – YM 2 AWAY	159
GRAPH 6.13	WM 1 AWAY – YM 3 AWAY	160
GRAPH 6.14	WM 2 AWAY – WM 3 AWAY	160
GRAPH 6.15	WM 2 AWAY – YMSPOT	160
GRAPH 6.16	WM 2 AWAY – YM 1 AWAY	160
GRAPH 6.17	WM 2 AWAY – YM 2 AWAY	160
GRAPH 6.18	WM 2 AWAY – YM 3 AWAY	160
GRAPH 6.19	WM 3 AWAY – YMSPOT	160
GRAPH 6.20	WM 3 AWAY – YM 1 AWAY	160
GRAPH 6.21	WM 3 AWAY – YM 2 AWAY	161
GRAPH 6.22	WM 3 AWAY – YM 3 AWAY	161
GRAPH 6.23	YMSPOT – YM 1 AWAY	161
GRAPH 6.24	YMSPOT – YM 2 AWAY	161
GRAPH 6.25	YMSPOT – YM 3 AWAY	161
GRAPH 6.26	YM 1 AWAY – YM 2 AWAY	161
GRAPH 6.27	YM 1 AWAY – YM 3 AWAY	161
GRAPH 6.28	YM 2 AWAY – YM 3 AWAY	161
GRAPH 6.29	WMSPOT – WM 1 AWAY	163
GRAPH 6.30	WMSPOT – WM 2 AWAY	163
GRAPH 6.31	WMSPOT – WM 3 AWAY	163
GRAPH 6.32	WMSPOT – YMSPOT	163
GRAPH 6.33	WMSPOT – YM 1 AWAY	163
GRAPH 6.34	WMSPOT – YM 2 AWAY	163
GRAPH 6.35	WMSPOT – YM 3 AWAY	164
GRAPH 6.36	WM 1 AWAY – WM 2 AWAY	164
GRAPH 6.37	WM 1 AWAY – WM 3 AWAY	164
GRAPH 6.38	WM 1 AWAY – YMSPOT	164
GRAPH 6.39	WM 1 AWAY – YM 1 AWAY	164
GRAPH 6.40	WM 1 AWAY – YM 2 AWAY	164
GRAPH 6.41	WM 1 AWAY – YM 3 AWAY	164
GRAPH 6.42	WM 2 AWAY – WM 3 AWAY	164
GRAPH 6.43	WM 2 AWAY – YMSPOT	165
GRAPH 6.44	WM 2 AWAY – YM 1 AWAY	165

GRAPH 6.45	WM 2 AWAY – YM 2 AWAY	165
GRAPH 6.46	WM 2 AWAY – YM 3 AWAY	165
GRAPH 6.47	WM 3 AWAY – YMSPOT	165
GRAPH 6.48	WM 3 AWAY – YM 1 AWAY	165
GRAPH 6.49	WM 3 AWAY – YM 2 AWAY	165
GRAPH 6.50	WM 3 AWAY – YM 3 AWAY	165
GRAPH 6.51	YMSPOT – YM 1 AWAY	166
GRAPH 6.52	YMSPOT – YM 2 AWAY	166
GRAPH 6.53	YMSPOT – YM 3 AWAY	166
GRAPH 6.54	YM 1 AWAY – YM 2 AWAY	166
GRAPH 6.55	YM 1 AWAY – YM 3 AWAY	166
GRAPH 6.56	YM 2 AWAY – YM 3 AWAY	166
GRAPH 6.57	WMSPOT – WM 1 AWAY	168
GRAPH 6.58	WMSPOT – WM 2 AWAY	168
GRAPH 6.59	WMSPOT – WM 3 AWAY	169
GRAPH 6.60	WMSPOT – YMSPOT	169
GRAPH 6.61	WMSPOT – YM 1 AWAY	169
GRAPH 6.62	WM 1 AWAY – YM 2 AWAY	169
GRAPH 6.63	WMSPOT – YM 3 AWAY	169
GRAPH 6.64	WM 1 AWAY – WM 2 AWAY	169
GRAPH 6.65	WM 1 AWAY – YMSPOT	169
GRAPH 6.66	WM 1 AWAY – YM 1 AWAY	169
GRAPH 6.67	WM 1 AWAY – YM 2 AWAY	170
GRAPH 6.68	WM 2 AWAY – WM 3 AWAY	170
GRAPH 6.69	WM 2 AWAY – YM 1 AWAY	170
GRAPH 6.70	WM 2 AWAY – YM 2 AWAY	170
GRAPH 6.71	WM 2 AWAY – YM 3 AWAY	170
GRAPH 6.72	WM 3 AWAY – YM 3 AWAY	170
GRAPH 6.73	YMSPOT – YM 1 AWAY	170
GRAPH 6.74	YMSPOT – YM 2 AWAY	170
GRAPH 6.75	YM 1 AWAY – YM 2 AWAY	171

## CHAPTER 7

GRAPH 7.1	WMSPOT – WM 1 AWAY	176
GRAPH 7.2	WMSPOT – YMSPOT	177
GRAPH 7.3	WMSPOT – YM 1 AWAY	177
GRAPH 7.4	WM 1 AWAY – WM 2 AWAY	177
GRAPH 7.5	WM 1 AWAY – YMSPOT	178
GRAPH 7.6	WM 1 AWAY – YM 1 AWAY	178
GRAPH 7.7	WM 1 AWAY – YM 2 AWAY	178
GRAPH 7.8	WM 2 AWAY – YM 1 AWAY	179
GRAPH 7.9	WM 2 AWAY – YM 2 AWAY	179
GRAPH 7.10	WM 3 AWAY – YM 3 AWAY	179
GRAPH 7.11	YMSPOT – YM 1 AWAY	180
GRAPH 7.12	YM 1 AWAY – YM 2 AWAY	180

## LIST OF TABLES

## CHAPTER 2

TABLE 2.1	WHITE MAIZE VOLUMES TRADED	13
TABLE 2.2	YELLOW MAIZE VOLUMES TRADED	13
TABLE 2.3	WHEAT VOLUMES TRADED	14
TABLE 2.4	SUNFLOWER VOLUMES TRADED	15
TABLE 2.5	SOYBEANS VOLUMES TRADED	17

## CHAPTER 5

TABLE 5.1	SUMMARY STATISTICS: WHITE MAIZE LOG RETURNS PRICES	72
TABLE 5.2	NORMALITY TESTS RESULTS: STANDARDISED RESIDUALS	77
TABLE 5.3	MONTHLY SEASONAL EFFECTS OF WHITE MAIZE LOG RETURNS PRICES	79
TABLE 5.4	SUMMARY STATISTICS: YELLOW MAIZE LOG RETURNS PRICES	80
TABLE 5.5	NORMALITY TESTS RESULTS: STANDARDISED RESIDUALS	84
TABLE 5.6	MONTHLY SEASONAL EFFECTS OF YELLOW MAIZE LOG RETURNS PRICES	85
TABLE 5.7	BEST FIVE MODELS CHOSEN BY THE AUTOMATIC MODELLING PROCESS	93
TABLE 5.8	BEST FIVE MODELS CHOSEN BY THE AUTOMATIC MODELLING PROCESS	98
TABLE 5.9	BEST FIVE PREDEFINED MODELS	104
TABLE 5.10	BEST FIVE PREDEFINED MODELS	105
TABLE 5.11	TESTS FOR IDENTIFIABLE SEASONALITY	106
TABLE 5.12	FINAL ESTIMATES OF THE FREE PARAMETERS FOR WHITE MAIZE	112
TABLE 5.13	FINAL ESTIMATES OF THE FREE PARAMETERS FOR YELLOW MAIZE	115
TABLE 5.14	SIGNIFICANCE ANALYSES OF THE FREE PARAMETERS FOR THE MAIZE DATA (PRELIMINARY TESTS)	118
TABLE 5.15	SIGNIFICANCE ANALYSES OF THE FREE PARAMETERS FOR THE MAIZE DATA (FINAL TESTS)	120
TABLE 5.16	FINAL ESTIMATES OF THE FREE PARAMETERS FOR THE MAIZE DATA (FINAL TESTS)	121

## CHAPTER 6

TABLE 6.1	UNIT ROOT TEST OF THE ORIGINAL SERIES – ADF	132
TABLE 6.2	UNIT ROOT TEST OF THE ORIGINAL SERIES – PP	132
TABLE 6.3	UNIT ROOT TEST OF THE FIRST DIFFERENCES – ADF	133
TABLE 6.4	UNIT ROOT TEST OF THE FIRST DIFFERENCES – PP	133
TABLE 6.5	JOHANSEN’S TEST FOR CO-INTEGRATION: LONG-RUN PAIRS	139
TABLE 6.6	PEARSON’S TEST FOR CORRELATION: LOG RETURNS	145
TABLE 6.7	PEARSON’S TEST FOR CORRELATION: RESIDUALS	146
TABLE 6.8	SPEARMAN’S RANK TEST FOR CORRELATION: LOG RETURNS	149
TABLE 6.9	SPEARMAN’S RANK TEST FOR CORRELATION: RESIDUALS	150
TABLE 6.10	BETA TEST: LOG RETURNS	154
TABLE 6.11	BETA TEST: RESIDUALS	154
TABLE 6.12	PRELIMINARY PAIRS PORTFOLIO	156
TABLE 6.13	CORRELATION COEFFICIENTS BETWEEN HENDERSON-CURVES	158
TABLE 6.14	CORRELATION COEFFICIENTS BETWEEN SEASONAL FACTORS SERIES	163
TABLE 6.15	PRELIMINARY PAIRS PORTFOLIO	167
TABLE 6.16	CORRELATION COEFFICIENTS BETWEEN SEASONAL FACTORS SERIES	168
TABLE 6.17	FINAL PAIRS PORTFOLIO	171

## CHAPTER 7

TABLE 7.1	FINAL RESULTS: ALL INDICATORS	188
TABLE 7.2	FINAL RESULTS: RSI	190
TABLE 7.3	FINAL RESULTS: THE STOCHASTIC OSCILLATOR	191
TABLE 7.4	FINAL RESULTS: BOLLINGER BANDS	193
TABLE 7.5	FINAL RESULTS: MACD	194
TABLE 7.6	FINAL RESULTS: MA	195

# CHAPTER 1

## Pairs Trading: An Introduction

*In evaluating people, you look for three qualities: integrity, intelligence, and energy.  
If you don't have the first, the other two will kill you.*

Warren Buffet  
(Hagstorm, 2005:102)

### 1.1 Introduction

There are several approaches that can be followed in managing an agricultural commodities portfolio. One approach, taken by several South African agricultural market players, is to simply regard the futures market as a means to hedge their positions against adverse price movements. These market players mainly include millers, animal feed producers and, in some instances, agricultural commodity farmers. Other South African agricultural commodity traders have a more speculative approach in their dealings on the futures market (Van Zyl, 2006).

Whatever the approach, South African agricultural commodity traders have been using pairs trading for years, some with great success, while others have lost fortunes on the market (Van Zyl, 2006). This thesis will attempt to demonstrate the use of pairs trading as a way to manage an agricultural commodities portfolio in South Africa, and contribute to the limited and incomplete literature regarding pairs trading in South Africa.

This chapter commences with a short literature review on pairs trading in section 1.2 below. This is followed by the problem statement in section 1.3. The research aims and objectives are then discussed in section 1.4. Thereafter, the methodology of this thesis is briefly explained in section 1.5 and the chapter outline is given in section 1.6. Finally, section 1.7 concludes the chapter with notes to the reader.

## 1.2 Pairs Trading: An Overview

### 1.2.1 Introduction

Pairs trading, and other market neutral strategies, have been around since the advent of listed markets. Previously, this trading strategy was almost exclusively utilised by large institutional investors as well as investment firms, such as hedge funds (Whistler, 2004:3). The reason for its success is that pairs trading captures gains independently from market performance because of its non-directional, relative-value investment approach (Ehrman, 2004:32). In order to elucidate the concept of pairs trading, section 1.2.2 gives a brief overview of the origin and history of pairs trading, which is followed by a short discussion on pairs trading in section 1.2.3.

### 1.2.2 A Brief Overview of the History of Pairs Trading

Jesse Livermore is the first trader recorded to have used pairs trading. Livermore made use of the basic principles of pairs trading in that he never analysed single shares on their own, but rather investigated the two top shares in a particular industry (Smitten, 2005:43). Because Livermore believed that a trend is only legitimate once both the two top shares move in tandem, he included “sister share” considerations, earning him the reputation of the original pairs trader (Ehrman, 2006:22).

Pairs trading (as primary investment strategy) was first used in the portfolios of high-net-worth individuals and institutional traders, who had the ability and the necessary resources to implement them successfully (Ehrman, 2006:19). With the advent of the hedge fund boom in the late sixties, pairs trading became a more prolific strategy (FTSE Index Company (FTSE), 2005:3). Thus, it was the hedge fund industry that developed pairs trading into a profitable market-neutral strategy, which could function independently from other investment strategies.

In the early nineties traders on Wall Street became interested in more quantitative methods of speculation, further increasing pairs trading’s popularity as a short-term speculation strategy. Today, pairs trading is among the proprietary “statistical arbitrage”

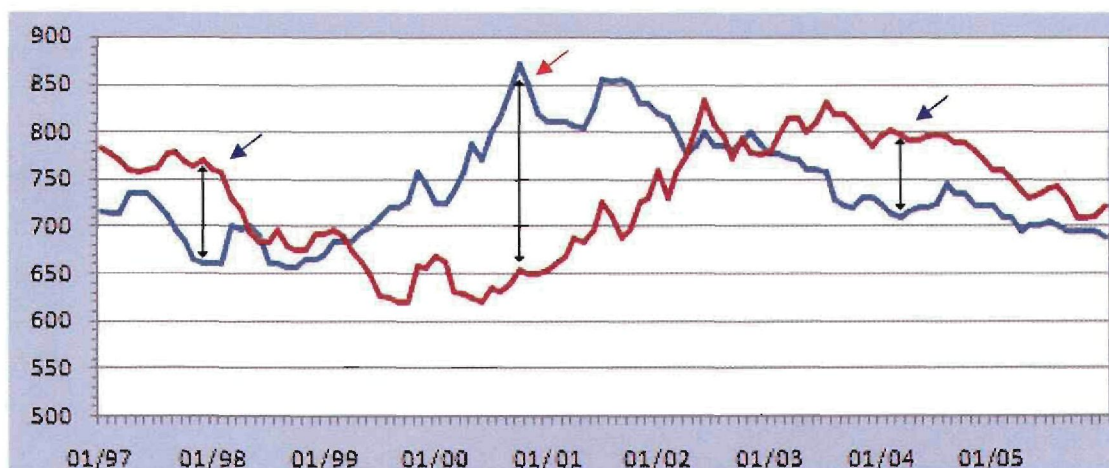
tools used by hedge funds as well as investment banks (Gatev, Goetzmann & Rouwenhorst, 1999:1).

### 1.2.3 What is Pairs Trading?

As mentioned in section 1.2.1, pairs trading is a market-neutral strategy that seeks to identify two financial instruments with similar characteristics.<sup>1</sup> For pairs trading to be effective, the relationship between these two instruments should be consistent over time. When the price relationship of the two instruments is trading outside their historical trading range, that is, the two instruments' price series are not in tandem anymore, the undervalued instrument is bought, while an equally large short position is instigated in the overvalued instrument.

The pairs trader will aim to enter the trade when the price gap between the two instruments is at its widest and close his positions when the price gap between the two instruments is at its widest in the opposite direction. However, should the gap between them grow wider over the duration of the long-short trade, the trader will lose money. Graph 1.1 below depicts the ideal entry and exit points for the pairs trade with the price of each instrument on the y-axis, and the date on the x-axis.

GRAPH 1.1 ENTRY AND EXIT POINTS IN A PAIRS TRADE



<sup>1</sup> Please note that the term 'financial instrument' refers to shares, commodities, and derivative instruments, among others.

The blue and red arrows indicate the entry and exit points where the gap between the two financial instruments' prices is the greatest. For the pairs trader to make the maximum amount of profit, he will typically enter the trade at the blue arrow, short selling the red instrument, and buying (going long) the blue instrument. He will thus earn a profit on both his long and short positions up until exiting the trade at the red arrow. After closing the position the trader can do one of two things: he can wait for the red instrument to reach a maximum gap from the blue instrument again before entering the market or, immediately take the opposite position, that is, buy the red instrument whilst short selling the blue instrument. If the trader chooses the second option, he will constantly have a position earning the maximum amount of profit.

Because pairs trading trades on the price gap on correlated instruments, a pairs portfolio will capture gains independent from market performance (Ehrman, 2006:2). It might happen that both instruments' prices go up, or that both fall. This is irrelevant as long as the long position increases more and faster than the short position, or the short position falls more and faster than the long position. Thus, it is possible for a trader to lose money on the long side whilst making money on the short side, or vice versa. The goal is, however, for the profits to exceed the losses over the trading period (Preston, 2005:44). The chosen pairs portfolio will thus remain profitable even if both shares rise or fall at the same time, as long as the relative gap between them narrows or widens.

For pairs trading to be effective, it is important to seek and identify two financial instruments with similar characteristics that are currently trading at a price relationship outside their historical trading range. When the two financial instruments' price relationship is distorted, the undervalued commodity is bought while short-selling the overvalued commodity, thereby maintaining market-neutrality (Ehrman, 2006:2).

Searching for financial instruments that will have the ideal relationship between them is often the most time consuming task in the pairs trading process. Most pairs traders look for co-movement between two financial instruments by testing the correlation between their prices (see for example, Gatev *et al.*, 1999; Castleman, 2003:38; McEwan, 2003:34; and Preston, 2005:40). Other pair traders look at this co-movement in terms of

the ratio between two correlated instruments (see for example, Ehrman, 2004:32; and Whistler, 2004:87).

### 1.3 Problem Statement

The problem regarding pairs trading in South Africa is three-fold. Firstly, although the international literature abounds in works on pairs trading and pairs trading is widely utilised by South African agricultural market participants, no literature exists specifically on pairs trading in agricultural commodities in South Africa. Secondly, there seems to be no written record of any analytical process to selecting agricultural commodity pairs (Van Zyl, 2006). Thirdly, no written record seems to exist on the performance of similar funds, nor could similar studies be found both in South Africa, or any other country.

All of the above problems are addressed by this thesis. Specific steps were devised and recorded regarding pairs trading in agricultural commodities in South Africa, and a detailed discussion is given on a comprehensive pairs selection process. A further contribution is made in identifying additional pairs selection tools.<sup>2</sup> This knowledge is useful for any agricultural commodities trader that would like to test his hand at pairs trading.

### 1.4 Research Aims and Objectives

The aim of this thesis is to provide the South African agricultural commodities trader with a comprehensive strategy to utilise pairs trading in successfully managing an agricultural commodities portfolio. Furthermore, this thesis aims to provide both the novice and the expert agricultural commodities trader with knowledge of this systematic process that will aid the selection and trading of agricultural commodities pairs in South Africa.

---

<sup>2</sup> Please note that only known methods were used in determining pairs. There are, however, no written records of seasonality and co-integration methods used in this way.

## 1.5 Methodology

The research aims of this study are achieved through basic research in the form of a literature review, and through performing empirical tests. The literature review considers all the elements that may influence the trader's trading decisions, including the history of agricultural commodities, fundamental analysis, and technical analysis. Empirical models, such as, Generalised Autoregressive Conditional Heteroskedasticity (GARCH), Autoregressive Integrated Moving Average (ARIMA) and Unobserved Component Model (UCM) are used to test for seasonality in the data in chapter 5. Tests for correlation, co-integration and Beta are also performed in chapter 6, in order to identify agricultural commodity pairs and validate them.

## 1.6 Chapter Outline

Chapter 2 looks into agricultural commodities with specific reference to those traded on the South African Futures Exchange (SAFEX). The chapter summarises the history of these agricultural commodities, after which the focus shifts to the factors that influence the prices of these commodities. This is done to elucidate the fundamental factors that will influence the agricultural commodities traders' trading decisions when entering into a pairs trade. It is to be noted that the inclusion of these factors serve to provide a more complete picture of the trading process.

When executing a pairs trading strategy in real-time, mathematical models built to model possible future pairs will include information regarding the fundamental factors that realise at the time. For the purpose of this thesis, more attention will however be given to technical analysis to determine entry and exit points for trades. The final section in chapter 2 briefly discusses the most prominent futures exchanges where agricultural commodities are traded.

Chapter 3 discusses technical analysis, and explores several market strength and moving average indicators. The comprehension of these indicators is essential to the success of the agricultural commodities pairs trader. Once pairs have been selected, these technical tools are used to identify entry and exit points.

Chapter 4 reviews the different models that are employed to test for seasonal patterns in the data in chapter 5. Two sets of models are utilised to this end: deterministic and stochastic models. The two deterministic models are introduced in the form of the Autoregressive Moving Average (ARMA-) and ARIMA-type models, and the general GARCH model, after which the stochastic UCM model is discussed.

Chapter 5 describes the tests for seasonal patterns in the data. Testing for seasonal patterns is done by making use of the models explained in chapter 4. The first test involves testing for seasonality by making use of several GARCH models. The second tests for the presence of seasonal patterns in the data by employing the X12 ARIMA-type test. Thereafter, testing for hidden seasonal patterns in the data is done by means of UCM. In order to establish whether these models will be effective in testing for seasonality in all the data series, preliminary tests are first performed. Tests that are successful in identifying seasonal patterns in the data are extended to the rest of the data. Finding seasonal patterns in the data is essential for identifying pairs during the pairs selection process.

Chapter 6 details the pairs selection process in two major sections. The first section discusses several pairs selection tools found in the literature and includes statistical measures such as co-integration, Pearson's correlation, Spearman's rank correlation and Beta. The second section reports on the results of the X12 tests described in chapter 5, and these results to further validate the pairs selected during the statistical tests. The final section of chapter 6 reports on the UCM results, which are crucial for the selection of pairs.

In Chapter 7 the results of test trading the final pairs portfolio is reported to ascertain whether it is possible to use this strategy successfully as an agricultural commodities pairs trader in South Africa. Prior to the trading process several important assumptions regarding the following are established: what entry and exit points consist of, the analysis tools used for identifying entry and exit points, the cost structure of the trades, and the process followed during trading. This chapter also draws conclusions regarding the profitability of this trading strategy.

Chapter 8 concludes the thesis with reference to the aims of this project. This is followed by a summary of all the factors to be considered when managing an agricultural commodities portfolio using a pairs trading strategy. Conclusions are also drawn with regard to the results of test trades performed in chapter 7. Finally, recommendations are made regarding future research.

## 1.7 Notes to the Reader

### 1.7.1 Publications

Three papers resulting from the research reported in this thesis have been presented on UCM.<sup>3</sup> The first paper, titled ‘Seasonality as an unobservable component in SAFEX agricultural market data’, was presented at the South African Finance Association’s (SAFA) annual conference in Cape Town, South Africa held from the 17th to 19th of January 2007.

The second paper, titled ‘Seasonality as an unobservable component in South African agricultural and stock market data’, was presented at the European Applied Business Research Conference in Padova, Italy held from the 4th to 7th of July 2007.

The third paper, titled ‘Seasonality as an unobservable component in South African agricultural market data’, was presented at the Economic Society of South Africa’s (ESSA) biannual conference held in Johannesburg, South Africa from the 10th to 12th of September 2007.

The article, titled ‘Seasonality as an unobservable component in South African agricultural and stock market data’ was also accepted for publication in the accredited International Business & Economics Research Journal (IBER), and is to be published in the March 2008 volume.

---

<sup>3</sup> Please refer to sections 4.5 and 5.5.

## 1.7.2 Miscellaneous

It is important to note the following:

- The term ‘research project’ (also only ‘project’) is used to refer to the research that this thesis reports on. In this way, ‘thesis’ refers to the physical text.
- Where the text refers to ‘agricultural commodities’, it includes white maize, yellow maize, sunflower, wheat and soybeans.
- Although ‘pairs trading’ and ‘spread trading’ are similar in the way they are executed, these concepts are not the same. According to Ehrman (2006:30), ‘spread trading’ constitutes the buying and short selling of the same amount of instruments, even though the monetary amount may differ. A pairs trade constitutes buying and short selling of the instruments to the same value, even though the quantity of instruments may differ.

Therefore, for the purpose of this thesis, ‘pairs trading’ constitutes buying a set amount of agricultural commodity futures contracts, whilst simultaneously short selling the same quantity of agricultural commodity futures contracts. The underlying commodities can be the same commodities with different times to maturity, different commodities with the same time to maturity, or different commodities with different times to maturity.

- The male gender is used throughout this thesis as the neutral gender.
- The list of acronyms folds out at the back of the thesis.

# CHAPTER 2

## Agricultural Commodities Trading: Fundamental Analysis

*It would be foolish, in forming our expectations, to attach great weight to matters that are very uncertain.*

John Maynard Keynes (1936:148)

### 2.1 Introduction

While agricultural commodity trading has been practised for millennia, the complexity of these transactions has changed drastically over time. Today, the agricultural commodities trader makes use of mainly technical analysis and fundamental analysis to assess the profitability of a given transaction. It is, therefore, imperative for the trader to have an in-depth understanding of these disciplines if he is to become, and remain, successful.

This chapter discusses fundamental analyses together with the history of the agricultural commodities currently being traded on SAFEX in South Africa. This knowledge will add to the understanding of the fundamentals surrounding agricultural prices, and of how these factors influence the trader's trading strategy.<sup>4</sup> This exposition of agricultural commodities will also promote the understanding of the relationships between various commodities. Because of similar characteristics among some commodities, it is often true that commodities compete, not only for available acreage, but also for market share, because of their similar uses. It is because of these similarities that some commodity prices have inherent trends.<sup>5</sup> This chapter commences with section 2.2 below in which the history of the agricultural commodities market is summarised. This is followed by a discussion on the factors that influence the prices of these commodities, in section 2.3. Finally, section 2.4 covers the highlights of several prominent agricultural commodities trading exchanges around the world.

---

<sup>4</sup> See chapter 7.

<sup>5</sup> Refer to chapter 5 where this knowledge is important for understanding the seasonal patterns found in the SAFEX data.

## 2.2 The History of the Commodities Traded on SAFEX

### 2.2.1 Introduction

In order to equip the reader with the knowledge needed to understand the inherent relationships between particular agricultural commodities a brief history of the agricultural commodities market is given. It is to be noted that agricultural commodities are used for both human and animal consumption, and that this distinction plays an important role in the determination of commodity prices. Corn (or maize)<sup>6</sup> are discussed first, followed by soybeans, wheat and lastly, sunflower.

The first part of this section discusses agricultural commodities with reference to their origin and how these markets have progressed to their current world status. The second part discusses the position of the South African agricultural commodities market with reference to the factors that influence the supply, demand and price for these commodities.

### 2.2.2 Maize

Maize is a large domesticated grass first cultivated in Mexico for human consumption purposes more than 5000 years ago. From Central America maize was taken to North America, and in the 15th century, the Spaniards and other Europeans took the plant to Europe (Salvador, 1997). Today, maize is planted in most countries, with total world production reaching 703 million metric tonnes during the 2006/07 season. The world's maize consumption for the same period was 721 million metric tonnes. The projected production for the 2007/08 season is 777 million metric tonnes, while projected consumption for this period is 770 million metric tonnes (South African Grain Information Service (SAGIS), 2007).

The bulk of today's maize production is no longer utilised mainly for human consumption. Since maize is relatively high in carbohydrates, it is used as the basic energy source in many animal feeding rations (Hinebaugh, 1985:7.10). In North

---

<sup>6</sup> Since 'corn' is better known as 'maize' in South Africa, the latter terminology is used throughout the thesis.

America, which is the largest consumer of maize, 75 percent of the country's maize production is used for animal feed (Geman, 2005:149). This maize, called 'dented corn', is used in the feed rations of cattle, hogs, poultry, and dairy cows (Hinebaugh, 1985:7.10).

Although yellow maize is predominantly used for animal feed, it is also used in the production of ethanol and vegetable oil. In the United States, more than 19 percent of all maize was used in the production of ethanol by the end of the 2006/07 crop year (Leibtag, 2008).

Another 8 percent of the United States' maize stock is used for the production of vegetable oil (Geman, 2005:150). This maize oil is an important ingredient in margarine, because of its low cholesterol content.

The only exception to the popular use for maize can be found in Southern Africa. Here, white maize is utilised as an important source of food for humans. South Africa is one of the largest producers of white maize, making this country an important role player in the white maize market. While South Africa also produces yellow maize, roughly 60 percent of South Africa's maize production consists of white maize (Geldenhuis, 2006). White maize is mainly used in the production of speciality food products. Between 1993 and 2003, the average annual production of white maize in South Africa was 4.3 million tons (Krugel, 2003:58). More recently, 6.62 million metric tonnes was produced in the 2005/2006 season, and an estimated 6.9 million metric tonnes for the 2006/2007 season (National Crop Estimate Committee (NCEC), 2007).

Both white and yellow maize futures contracts are the best represented agricultural commodities on SAFEX. These contracts are also active for longer periods than the other agricultural commodities traded on SAFEX. The July contracts for both white and yellow maize are often active for periods of fifteen months and more (SAFEX, 2007). Tables 2.1 and 2.2 below, display the opening price, the low, the high, and the closing price for the day for both white and yellow maize. The final row in each table shows the amount of contracts traded on the particular day.

TABLE 2.1 WHITE MAIZE VOLUMES TRADED

<b>Open</b>	July 2007	1259	1269	1294	1308	1280	1236	1263	2 Contracts away
<b>Low</b>		1253	1271	1301	1265	1235	1235	1298	
<b>High</b>		1277	1295	1330	1300	1253	1272	1308	
<b>Close</b>		1269	1294	1308	1280	1236	1263	1308	
<b>OI</b>		7407	7422	7539	7669	7650	7724	7736	
<b>Dates</b>		<i>06/12/28</i>	<i>06/12/29</i>	<i>07/01/02</i>	<i>07/01/03</i>	<i>07/01/04</i>	<i>07/01/05</i>	<i>07/01/08</i>	

TABLE 2.2 YELLOW MAIZE VOLUMES TRADED

<b>Open</b>	July 2007	1265	1257	1284	1291	1261	1238	1253	2 Contracts away
<b>Low</b>		1255	1275	1290	1255	1216	1237	1280	
<b>High</b>		1275	1290	1315	1280	1239	1255	1298	
<b>Close</b>		1257	1284	1291	1261	1238	1253	1298	
<b>OI</b>		4120	4106	4104	4106	4139	4142	4182	
<b>Dates</b>		<i>06/12/28</i>	<i>06/12/29</i>	<i>07/01/02</i>	<i>07/01/03</i>	<i>07/01/04</i>	<i>07/01/05</i>	<i>07/01/08</i>	

While Table 2.1 shows the trading data for the July 2007 contract on white maize from the 28th of December 2006 to the 8th of January 2007, Table 2.2 displays the trading data for the yellow maize data for the same period. White maize contracts are far more liquid than yellow maize contracts. In December 2006, an average of 7133 July white maize contracts were traded per day, while an average of only 4105 July yellow maize contracts were traded per day over the same period (SAFEX, 2007).

### 2.2.3 Wheat

After the ice age, *circa* 8000 B.C, the persisting drought drove people in Syria to depend increasingly on wild grass seeds (Geman, 2005:150). They began to cultivate rye and chickpeas, and later einkorn and emmer, two ancestors of wheat. Wild einkorn grass (containing the identical genetic fingerprint of modern domesticated wheat) soon became the staple food of the time (Geman, 2005:150). By 1529, the Spanish brought wheat to Mexico, where it later spread to America and Canada (Heiser, 1981:84). Since its early use in Syria, wheat has been widely used as a food source for both humans and animals.

Today, although still used for animal feed, wheat is mainly used as a food source for human consumption. In this regard, wheat is processed to manufacture foods such as sweet goods, bread and pastry products (Hinebaugh, 1985:6.4). Although wheat is widely used in the production of these products, it competes with rice as a good source of carbohydrates, thus rendering its price dependent on annual rice harvests.

The major wheat producing countries include the United States, China, and the Russian Federation. In the 2006/2007 season, these countries contributed to world wheat production of 593 metric tonnes. This production was outstripped by consumption at 618 metric tonnes for the same period (SAGIS, 2007). Similar to maize, wheat futures contracts are well represented on SAFEX. However, wheat contracts are not traded for such long periods as maize. These contracts are never active for periods exceeding twelve months for any of the contract months (SAFEX, 2007). Table 2.3 below, displays the opening price, the low, the high, and the closing price for the day for wheat.

TABLE 2.3 WHEAT VOLUMES TRADED

<b>Open</b>	July 2007	1776	1796	1825	1833	1820	1807	1826	2 Contracts away
<b>Low</b>		1770						1855	
<b>High</b>		1770						1868	
<b>Close</b>		1796	1825	1833	1820	1807	1826	1855	
<b>OI</b>		136	136	136	136	136	136	158	
<b>Dates</b>		06/12/28	06/12/29	07/01/02	07/01/03	07/01/04	07/01/05	07/01/08	

From Table 2.3 above, it is clear that the wheat contracts are illiquid. The fact that the volume for the period of the 28th of December to the 5th of January remains constant on 136, is an indication that there were no deals placed on those days. Although wheat contracts are not as liquid as maize contracts, certain contract months can reach up to 13000 contracts traded per day (SAFEX, 2007).

### 2.2.4 Sunflower

Unlike most other crop species, sunflower originated in North America. Sunflower was domesticated by western Native American tribes more than 3000 years ago (Putnam *et al.*, 1990). Europe was only exposed to this new crop during the exploration of the Americas a few centuries ago. Its spread through Europe was slow and it was only

harvested for commercial purposes by Russia in 1860. It was only after the discovery of the high oil producing cultivars, in Russia, that the United States renewed their interest in sunflower. Production of sunflower subsequently increased dramatically in the Great Plains states (in the United States) as marketers found new niches for the seeds as human snack food, oil crop, animal feed, and birdseed (Putnam *et al.*, 1990).

Like wheat and maize, sunflower can be found in a large number of countries. Although sunflower is grown in so many countries, there are only three major sunflower producing areas: the Russian Federation, the EU-25, and Argentina (United States Department of Agriculture (USDA), 2007).<sup>7</sup>

Sunflower is mainly harvested for its high oil content with commercially available sunflower varieties containing up to 50 percent oil in the seed (Heiser, 1981:186). Sunflower oil is generally considered first-rate oil because of its high level of unsaturated fatty acids, light colour, bland flavour, and high smoke points. Therefore, Sunflower oil is primarily used in margarine, as well as salad and cooking oil (Putnam *et al.*, 1990). Despite its various uses, sunflower is not harvested in the same quantities as wheat and maize are, and makes up only a small portion of the world's agricultural commodity production. Sunflower seed production only amounted to 30 million metric tonnes for the 2006/07 season (USDA, 2007).

Similarly to wheat, sunflower futures contracts are also not active for periods exceeding twelve months for any of its contract months (SAFEX, 2007). Table 2.4 below, displays the opening price, the low, the high, and the closing price for the day for sunflower.

TABLE 2.4 SUNFLOWER VOLUMES TRADED

<b>Open</b>	July 2007	2480	2480	2470	2465	2445	2440	2471	2 Contracts away
<b>Low</b>		2480	2470	2465	2445	2440	2460	2480	
<b>High</b>		2480	2470	2465	2445	2440	2462	2514	
<b>Close</b>		2480	2470	2465	2445	2440	2471	2514	
<b>OI</b>		126	126	126	126	130	132	136	
<b>Dates</b>		06/12/28	06/12/29	07/01/02	07/01/03	07/01/04	07/01/05	07/01/08	

<sup>7</sup> EU-25 includes Austria, Belgium, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

Like the wheat data, the sunflower data also reveal illiquidity for the period of the 28th of December to the 5th of January. Despite liquidity levels picking up as the contract nears expiry, contracts traded per day never reaches 3000 (SAFEX, 2007).

### 2.2.5 Soybeans

Soybeans come from the soybean pods that are found in the soy plant. Soybean pods were first cultivated in China more than 5000 years ago and first appeared in France in the 18th century and later in the United States at the end of the 19th century (Heiser, 1981:142). Soybeans were primarily used as animal feed, and quickly became popular because of the similarity to the maize culture.

Apart from its popularity as replacement for maize products, soybeans are also planted to restore the nitrogen in the soil that maize and other crops take out. This led to its use as a rotation crop with maize (Hinebaugh, 1985:8.1). Today, the United States produces 45 percent of world's soybeans, with Brazil and Argentina producing 36 percent (Geman, 2005:146). World production for the 2006/2007 season reached 226.78 million metric tonnes.

Although soybeans are mainly produced by only three countries, this commodity is used around the world. Soybeans are crushed for the production of soybean meal and oil. Soybean meal makes up 80 percent of the total content of the soybean and oil 20 percent. Since soybean meal is protein rich, it replaced fishmeal as the main source of protein in animal feeds. Furthermore, soybean meal accounts for two-thirds of the total world production of protein meals for humans (Hinebaugh, 1985:8.1). In terms of edible oil, soybeans account for 13 percent of world production (USDA, 2007). Because of its use as cooking oil, soybean oil competes with groundnut, canola, palm, and sunflower oil (Geman, 2005:149). Apart from its use as cooking oil, lecithin, an extract of soybean oil, is also used in many food preparations as an emulsifier.

Like wheat and sunflower, soybean futures contracts are also not active for periods exceeding twelve months for any of its contract months (SAFEX, 2007). Table 2.5 below, displays the opening price, the low, the high, and the closing price for the day for soybean futures contracts.

TABLE 2.5 SOYBEANS VOLUMES TRADED

<b>Open</b>	July 2007	2160	2190	2185	2185	2155	2195	2195	2 Contracts away
<b>Low</b>				2195		2155		2230	
<b>High</b>				2195		2205		2245	
<b>Close</b>		2190	2185	2185	2155	2195	2195	2230	
<b>OI</b>		138	138	138	138	140	140	138	
<b>Dates</b>		06/12/28	06/12/29	07/01/02	07/01/03	07/01/04	07/01/05	07/01/08	

Similarly to the wheat and sunflower data, soybean futures contracts also reveal illiquidity for the period of the 28th of December to the 5th of January. This agricultural commodity is the least liquid and contracts traded per day never reach 2000 for any of its contract months (SAFEX, 2007).

## 2.2.6 Conclusion

Although most agricultural commodities cultivated today, have served as a source of food for both people and animals for millennia, some of these commodities have been utilised for different purposes. So, for example, it is possible to manufacture margarine from a variety of commodities including, maize, sunflower and canola. It is for this reason that some agricultural commodities prices show similarities. Section 2.3 below explores these inherent relationships in more depth.

## 2.3 Agricultural Commodity Futures Prices

### 2.3.1 Introduction

Fundamental analysis focuses on the factors that influence the supply and demand of the commodity.<sup>8</sup> Any increase in the supply of the commodity will decrease the price of this commodity and therefore, also reduce the price of any derivative contracts linked to this commodity. Similarly, any increase in the demand of the commodity will increase the price of any derivative contracts linked to this commodity, while any decrease in the demand of the commodity will lead to a decrease in the derivative contract price (Bernstein, 2000:148). Conversely, any decrease in the supply of the underlying

<sup>8</sup> 'Commodities' is used to refer to agricultural commodities

commodity will increase the price of the underlying commodity that will lead to an increase in the price of the derivative contract.

Having knowledge of the factors that influence the supply and demand for agricultural commodities will empower the agricultural commodities trader to have a better insight into what a specific futures contract should be worth. Section 2.3.2 below discusses these factors. This is followed by a discussion on other factors that influence the price of a futures contract in section 2.3.3. Section 2.3.4 discusses futures exchanges in terms of the products they trade as an indication of world agricultural commodities trade.

## 2.3.2 The Supply and Demand of Commodities

The fundamental factors determining the price of agricultural commodities such as maize in South Africa is: the supply and demand at international level, domestic supply, demand and stock levels, and the rand-dollar exchange rate, since this directly affects the import and export parity price (Geysers & Cutts, 2007:296).

The demand side of agricultural commodities is determined by population growth, the availability of substitute products, and the stock-to-use ratio. On the supply side, available acreage, technology, the production of the commodity in the current year, imports from other countries, the surplus stock from the previous year, this is known as carry-in, and the stock-to-use ratio play a role in determining the price of the commodity (Geldenhuys, 2006).<sup>9</sup> However, it should be noted that it is not only the supply and demand of the commodity itself that will determine its price, but also the supply and demand factors of these commodities' substitute and complementary products.

The demand for certain commodities might change over time as new methods are found to utilise these products. The use of maize sweeteners as a healthier substitute for sugar cane sweeteners, for example, led to an increase in the demand for maize in the early and mid-eighties (Hinebaugh, 1985:7.13). The replacement of fishmeal with soybean meal as a source of protein is another example.

---

<sup>9</sup> The stock-to-use ratio is calculated by dividing the current year's ending stock by the current year's use.

Artificial demand can also be created as a result of the relationship between raw and refined products. Soybean meal and oil are good examples of this phenomenon. Soybean meal makes up 75 to 80 percent of the soybean, and is used for animal feed, directly competing with maize (Geman, 2005:148). Soybean meal prices are also influenced by the availability of meal from crushing operations, the price of fishmeal, the price of maize and the size of livestock herds. Soybean oil is used for cooking, and competes with canola, sunflower, groundnut and palm oil (Geman, 2005:149).

If the demand for soybean oil is low while the demand for soybean meal is high, processors will continue to crush soybeans, sell the meal, and store the oil until the price for soybean oil rises to acceptable levels (Geman, 2005:149). An artificial demand for the oil is created by storing the soybean oil, a necessary product of the production of soybean meal.

Another good example of this phenomenon is the relationship between white and yellow maize in South Africa. Because white maize can be used as a substitute for yellow maize, the producers of most animal feed will start to use white maize if yellow maize becomes too expensive (Van Zyl, 2006).<sup>10</sup>

Technology also plays an important role in the production of commodities. Technology, in this sense, mainly refers to the use of nitrogenous fertilisers, but also includes the farming implements used for planting and harvesting. Since technology enables farmers to produce their commodities at lower costs, it also affects carry-in, another important supply side factor. Carry-in from previous years will be high if conditions for planting were favourable, and the price of the commodity was high enough to encourage farmers to produce the specific commodity. Naturally, carry-in will be low if conditions in preceding seasons discouraged farmers from producing the commodity (Geman, 2005:143). In turn, planting conditions are mainly driven by weather patterns.

Weather patterns do not only impact on local production, but also on imports from other countries, as well as the local surplus stock left from the previous years. Higher levels

---

<sup>10</sup> Note that only some feed producers will use white maize in their products. Feed for chickens and pigs is usually not made with white maize, even if the price of white maize is below that of yellow maize (Van Zyl, 2006).

of rainfall are mostly associated with increased supply, thus lower commodity prices, while lower levels of rainfall will have the opposite effect. Both realised, as well as estimate rainfall, have an impact on commodity prices (Kleinman, 2001:114).

Weather patterns are predicted by phenomena, such as the Southern Oscillation, which is a widespread inter-annual oscillation in sea-level pressure between one region near northern Australia and one in the central Pacific Ocean. Even though the Southern Oscillation Index (SOI) is computed on a daily basis, these values do not convey much in terms of useful information about the state of the climate. The values are thus converted to monthly or seasonal values, and then used effectively as indicator of the related phenomena, El Niño-Southern Oscillation (ENSO) and anti-ENSO (Australian Bureau Meteorology (ABM), 2006).

ENSO is used colloquially to describe the whole cluster of changes associated with an 'El Niño' event. This includes, among others, changes in rainfall, and atmospheric pressure. The warm phase of ENSO is related to El Niño, which is the widespread warming of the upper ocean in the tropical eastern Pacific over a period of five months or more (Hansen *et al.*, 1999:93). This leads to increased cloudiness in the central tropical Pacific Ocean, weaker than normal easterly winds, and low or negative values of the SOI. This phase of ENSO is usually associated with drier conditions (ABM, 2006).

The cold phase of ENSO is associated with La Niña, and refers to the extensive cooling of the central and eastern Pacific Ocean (Wang, Zhang, Cole, & Chavez, 1999). During this anti-ENSO phase, increased cloudiness is generated over tropical Australia, Papua New-Guinea, and Indonesia (ABM, 2006). This phase usually lead to stronger than normal easterly winds across the Pacific Ocean and high or positive values of the SOI, and is usually associated with the increased probability of wetter conditions (ABM, 2006).

El Niño and La Niña are particularly important for the supply of grains across the world. In South Africa and Australia, for example, dry conditions will persist in an El Niño year, while good rainfall will be experienced in Mid-west America (Van Zyl, 2006).

Conversely wet conditions are associated with a La Niña year for countries in the Southern hemisphere (Hoerling, Kumar & Zhong, 1997:741). Therefore, it is common to experience high grain prices during an El Niño year and low prices during La Niña years for countries such as South Africa.

The indices discussed above are often used to explain commodity prices (Hansen *et al.*, 1999; Martin *et al.*, 2000), thus indicating that weather patterns do play an important role in determining the price of commodities. However, for good yields it is not only important to have favourable circumstances during the growth phase of the plants, but also to have good soil moisture reserves before planting (Rossouw, 2006).

From the discussion above, it is clear that the supply of and the demand for agricultural commodities, as well as their complements and substitutes are mainly responsible for the changes in their prices. These factors are, however, not the only deciding factors regarding commodity prices, but future expectations also play a role in determining commodity prices; therefore, it is important to discuss derivative contract pricing briefly.

## 2.3.3 Other Factors Influencing Derivative Contract Pricing

### 2.3.3.1 Introduction

The futures contract price ultimately reflects the price at which buyers, and sellers of the contract, are willing to buy or sell the underlying commodity at a future date. Futures contract prices reflect the supply and demand of the underlying commodity and the price at which this commodity will be traded at on a future date.<sup>11</sup> This futures contract price reflects all currently available market data and changes as new market information becomes available (Krugel, 2003:77).

The remainder of this section attempts to shed some light on the determinants of commodity futures contract prices. The discussion commences with the relationship between the cash and futures prices of commodities in section 2.3.3.2, followed by a

---

<sup>11</sup> See section 2.3.1 above for the factors that influence the supply and demand for commodities.

short discussion on contango and backwardation markets in section 2.3.3.3. Section 2.3.3.4 discusses the factors that influence the pricing of futures contracts.

### 2.3.3.2 The Basis

According to Strong (2002:420), total basis is the difference between a futures price for a commodity and the cash price of the commodity at a specific location. Total basis can be divided into carry basis and value basis. Carry basis is a theoretical future price minus the spot price of the underlying asset, and is equal to the net cost of carry. Value basis is the difference between the theoretical future price and its market price (Watsham, 1998:88). The basis can be calculated as follows (Kolb, 1997:63):

Basis = Current cash price – Futures price.

Because the cash price of a commodity differs from one location to the next, it follows that the basis for that commodity will also differ accordingly. The reason for this difference in price is brought about by the difference in storage and transportation costs. Basis risk, therefore, refers to the instability in the basis caused by the difference in the cash price of a commodity from one location to the next, for reasons other than the costs of storage and transportation (Kleinman, 2001:21).

The basis can have a positive or negative value based on the relationship of the cash price and the futures price. In the case where the futures price is higher than the current cash price, the basis will be negative, conversely, if the current cash price is higher than the futures price, the basis will be positive (Kolb, 1997:64). A negative basis value is called a contango market while a positive basis value refers to a backwardation market (Strong, 2001:419, 421). In theory, the basis ought to be zero on the futures delivery date, since the cash and futures prices are equal (Kolb, 1997:65).

### 2.3.3.3 Contango and Backwardation Markets

A contango, or normal market, exists where the prices for nearby futures contracts are lower than the prices for more distant futures contracts<sup>12</sup>. Since the cash price of the

---

<sup>12</sup> Nearby futures contracts have an expiration date closer in the future while distant futures expire further into the future.

underlying commodity is lower than the futures price, the basis will increase from its negative value until it is zero at expiration (Kolb, 1997:65).

A backwardation, or inverted market, exists where the prices of nearby futures contracts exceed the prices for more distant futures contracts. Because the cash price of the underlying commodity is higher than the futures price in this case, the basis will decrease from its positive value until it is zero at expiration (Kolb, 1997:65).

Apart from the relationship between the futures and cash price, there are also relationships between futures prices called spreads. Strong (2002) identifies three different types of spreads: inter-commodity spreads, inter-market spreads and intra-commodity spreads.

Inter-commodity spreads require holding a long and short position in two related commodities (Strong, 2002:217). An inter-market spread involves taking opposite positions in two different markets. In this case, a speculator may buy a commodity on the cash market (at a price lower than the futures price) to profit from selling it on the futures market (Strong, 2002:218). An intra-commodity spread requires taking different positions in different delivery months for the same commodity (Strong, 2002:218).<sup>13</sup>

The next section discusses the factors that influence the pricing of a futures contract. With this information at his disposal, it will be easier for the commodities trader to manage a pairs portfolio more effectively.

#### 2.3.3.4 Factors Influencing the Pricing of Futures Contracts

Consumption commodities such as agricultural commodities are not held for their financial return only, but also for the role they play when they are consumed in the production process (Watsham, 1998:93). Therefore, it is not possible to derive the futures price of an agricultural commodity purely on the availability of the underlying commodity. This makes it impossible to rely on the arbitrage process to ensure that

---

<sup>13</sup> Intra-commodity spreads are also popular in pairs trading strategies where the two contracts form a pair.

commodity futures contracts trade below  $P_t + C$ ; where  $P_t$  is the price of the commodity and  $C$  is the net cost of carry (Watsham, 1998:86).

However, if the potential consumer is not interested in the immediate consumption of the commodity, he might sell his future contract short and buy the commodity with the objective of delivering the commodity. This being the case, it is to be expected that the activities of commodity consumers will ensure that the futures price does not exceed  $P_t + C$ . Such a commodity futures contract will be priced below  $P_t + C$  to the degree that it is convenient for the commodity holder to have the commodity in his possession, in order to facilitate the production process. The agricultural commodity futures price can thus be given as (Watsham, 1998:94):

$$F = P + (C - Y), \quad (2.1)$$

where  $Y$  is the monetary value given to the convenience yield.<sup>14</sup> It is thus clear that the pricing of consumption commodities futures contracts differs from that of non-consumption commodities.

### 2.3.4 Conclusion

For the commodity pairs trader to be successful, it is imperative that he is well aware of all the fundamental factors that play a role in the determination of agricultural commodity prices. This includes knowledge of weather patterns, new technology, complementary and substitute products, the uses of these products, and pricing models. With this information at hand, it is possible for the experienced commodities trader to make an informed guess of what a specific commodity should be worth at a given time. Once the trader is aware of what the price for a commodity should be, he will automatically be able to determine whether a commodity is over valued, under valued, or on par.

---

<sup>14</sup> The convenience yield is an adjustment to the cost of carry in the non-arbitrage pricing formula for forward prices in markets with trading constraints.

To further elucidate trading activities regarding agricultural commodities, the next section briefly discusses the prominent futures exchanges around the world with reference to the commodities traded on them.

## 2.4 Futures Exchanges

### 2.4.1 Introduction

Once the pairs trader is comfortable with the agricultural commodity pairs portfolio in a single market, it might be a profitable notion to extend his portfolio to include commodity pairs in different markets. For example, it would be possible to find a commodity pair on the Chicago Board of Trade (CBOT) between lean hogs and maize futures contracts, or to pair white maize futures contracts on SAFEX with maize futures contracts on CBOT.

For the pairs trader to capture profits on these international deals, it is imperative that these commodities share a statistical relationship that allows them to be paired. Consequently, it is necessary for the pairs trader to identify the types of agricultural commodities available on other markets, and assess their relationships with each other. In order to explore these possibilities, a short discussion about the most prominent commodity markets, and the commodities traded on these markets, is given below.

### 2.4.2 CBOT

The CBOT, one of the oldest derivatives exchanges, was established in 1848 (Watsham, 1998:7). In 1851, CBOT recorded the first forward contract on maize. In 1865 they formalised grain trading by developing futures contracts, and in 1936 soybean futures contracts were launched. CBOT recorded its highest yearly volume traded in history, trading more than 674 million contracts in 2005 (CBOT, 2006a). This exchange currently trades agricultural derivative contracts on maize, wheat, soybean meal, soybean oil, ethanol, oats and rough rice. Derivative instruments are also available on interest rates and equities (CBOT, 2006a).

### 2.4.3 Chicago Mercantile Exchange (CME)

The CME was founded as the Chicago Butter and Egg Board in 1898, and evolved into the CME in 1919. Although CME deals in foreign exchange, equities, interest rate and weather derivatives, it deals in a range of agricultural commodities. These include classes 3 and 4 milk, feeder cattle, frozen pork bellies, lean hogs, live cattle as well as lumber (CME, 2006).

On the 17th of October 2006, the CME and CBOT announced that they had signed a definitive agreement to merge (CBOT, 2006b).

### 2.4.4 New York Board of Trade (NYBOT)

The NYBOT was founded as the New York Cotton Exchange (NYCE) in 1870. It changed its name to the Coffee Exchange of the City of New York in 1882, dealing in cotton and coffee futures respectively. In 1914, the Coffee Exchange added sugar futures and became the Coffee and Sugar Exchange in 1916. The New York Cocoa Exchange commenced with operations in 1925 and merged with the Coffee and Sugar Exchange in 1979 to form the Coffee, Sugar and Cocoa Exchange (CSCE) (NYBOT, 2006).

In 1998, the CSCE and NYCE formed the Board of Trade of the City of New York, as a parent company, and completed the merger process to form the New York Board of Trade (NYBOT) in June 2004 (NYBOT, 2006). Currently NYBOT deals in derivative contracts on soft commodities, such as cocoa, coffee, sugar, cotton and wood pulp (NYBOT, 2006).

### 2.4.5 Euronext.liffe

The London International Financial Futures and Options Exchange (LIFFE) was originally established as a financial futures and options exchange in 1982, after the removal of foreign exchange controls in the United Kingdom. In 1992, LIFFE merged with the London Traded Options Market (LTOM), and in 1996 with the London

Commodity Exchange (LCE), thereby adding a range of equity options, and soft and agricultural commodity contracts to its financial portfolio (Euronext, 2006).

LIFFE was acquired by the Euronext group in the beginning of 2002. Today, Euronext.liffe is Europe's leading cross-border exchange; integrating, trading, and clearing operations on both regulated and non-regulated markets for cash as well as derivatives products (Euronext, 2006). Apart from its equities division, Euronext.liffe also offers derivative contracts on cocoa, Robusta coffee, maize, rapeseed, raw and white sugar, as well as feed and milling wheat.

#### 2.4.6 Tokyo Grain Exchange (TGE)

The oldest futures exchange in Japan, established in 1730, came as a response from the feudal government of Tokugawa to the rice merchants' requests to authorise rice transactions at the Dojima Rice Market in Osaka (TGE, 2006).

The TGE originated from the Kakigaracho Rice Trading Exchange established in 1874 by a group of the Chugai Shoko Kaisha. The exchange later changed its name to the Tokyo Rice Trading Exchange, and in 1908, to the Tokyo Rice and Commodities Exchange. It was only when it opened again after the war in 1952, that it became known as the TGE (TGE, 2006).

After merging with the Tokyo Sugar Exchange in 1993, the Hokkaido Grain Exchange in 1995, and the Yokohama Commodity Exchange in 2006, the TGE extended its range of products. Today, futures contracts on the TGE include maize, soybean meal, soybeans, non-genetically modified soybeans, Azuki, Arabica coffee, Robusta coffee, raw sugar, raw silk, and vegetables (TGE, 2006).

#### 2.4.7 Sydney Futures Exchange (SFE)

The second big futures exchange operating in the Asia Pacific region is the SFE. The SFE started out as the Sydney Greasy Wool Futures Exchange in 1960, and is currently the tenth largest financial futures and options exchange in the world by volume turnover and the second largest in the Asia Pacific region (SFE, 2006).

The SFE currently offers futures contracts on grains, such as milling wheat, feed wheat, feed barley, canola and sorghum. Option contracts are also available on milling wheat, feed barley, canola and sorghum. Other commodity products include greasy wool futures, greasy wool options, fine wool futures, broad wool futures and cattle futures (SFE, 2006).

#### 2.4.8 South African Futures Exchange (SAFEX)

The SAFEX is one of the youngest futures exchanges and was started in 1988 by twenty-one banks and financial institutions. In 1990, the Safex Clearing Company (Safcom) took over as clearing company from Rand Merchant Bank (RMB). By 1995 the Safex Agricultural Markets Division (AMD) opened, listing its first commodity on the exchange in the form of a physical settled beef contract. This issue was shortly followed by a potato contract (SAFEX, 2006).

Both the physically settled beef and potato contracts were de-listed due to inactivity. It was rather the white and yellow maize contracts listed in 1996 that were responsible for the growth in contract volumes traded on the exchange. Today, SAFEX forms part of the Johannesburg Stock Exchange (JSE) after it was bought in 2001 (SAFEX, 2006).

SAFEX offers options and futures contracts on white maize, yellow maize, wheat, sunflower seeds and soybeans. Although other contract months exist, the December, March, July and September contracts are the most popular and most frequently traded on all the commodities (SAFEX, 2006).

#### 2.4.9 Conclusion

Although the range of products varies between the exchanges discussed above, it is clear that the majority of the products do coincide. This allows the trader to search for pairs between similar products over different exchanges, in addition to other 'cross-commodity' pairs. A more detailed discussion about different types of pairs will be given in chapter 5.

## 2.5 Summary

Even though technical analysis plays an important role in agricultural commodities trading, fundamental analysis will never be obsolete. A good understanding of the fundamentals of commodity futures will always enable the trader to understand the relationships between various commodities, and therefore, the relationships between various pairs. This knowledge will also help the agricultural commodities trader to know and understand how a futures price is determined, and what influence certain events might have on it.

For the pairs trader to be successful, it is therefore essential that he is well informed about all the fundamental factors that have an impact on the supply and demand of the commodities he intends to trade. This knowledge will typically include information about global weather patterns, the supply and demand for substitute and complimentary products as well as information about developments in the various derivatives markets for these commodities.

In addition to the fundamental analysis discussed above, the commodity pairs trader must also be well versed in the use of technical trading tools. Chapter 3 below will discuss the most commonly used technical analysis tools used for trading agricultural commodities. These tools, together with the fundamental analysis discussed here, and the exposition of the seasonal patterns in chapter 5, will aid the trader to successfully manage an agricultural commodities pairs portfolio, which will be discussed in chapter 7.

# CHAPTER 3

## Technical Analysis

*Technical analysis and fundamental analysis should always work hand-in-hand, as indeed should complex analysis and common sense.*

(McHattie, 1996:98)

### 3.1 Introduction

While technical analysis is the study of market behaviour, the study of market behaviour is greatly the study of human psychology. This is indicated by the observation that history tends to repeat itself, and that market trends usually persist in the future, see for example Murphy (1986:4), Pring (2002:3), Chande (2001:16). Since these trends exist in the market, it is critical that the pairs trader is aware of these market trends, and that he is able to identify, and predict them with reasonable accuracy.

Ehrman (2006:100) distinguishes between three groups of technical indicators most relevant to pairs trading. The first group falls under the market strength indicators, while the second group is known as moving average indicators. The third major technical indicator is considered volume as indicator.<sup>15</sup> Because this approach to market analysis is also key to the success of a pairs trader, chapter 3 explores technical analysis as a preamble to chapter 7. Therefore, sections 3.2 and 3.3 briefly discuss some of the most frequently used technical analysis tools, in order to complement the fundamental analysis covered in chapter 2.

### 3.2 Market Strength Indicators

#### 3.2.1 Introduction

Although there is a wide range of market strength indicators, such as William's %R, the Relative Strength Index (RSI), and the Stochastic oscillator, only the two most commonly used indicators are discussed here. These indicators give the trader an idea of

---

<sup>15</sup> Since each leg of the pair has its own volume, but no common volume (of contracts traded for the day) exists for the pair, volume will not be discussed here as a technical analysis tool.

the speed with which the market is moving in a particular direction. Similarly, it is possible to establish when the market is slowing down its movement and when a possible turnaround can be expected. Before discussing these indicators, it is necessary that the key concepts: overbought, oversold, and momentum are defined.

### 3.2.2 Overbought and Oversold

In the case where a market or agricultural commodity is not overbought or oversold, market forces keep the number of buyers and sellers in equilibrium. A market or agricultural commodity is considered overbought when the recent number of buyers has been disproportionately high compared to the number of sellers. When such a condition prevails, the market price will fall as the number of sellers increases in relation to the number of buyers (Meyers, 1994:299).

The opposite is true for an oversold condition. When an oversold condition occurs, the recent number of sellers has been disproportionately high compared to the number of buyers (Meyers, 1994:299). In the presence of an oversold condition, the market price will rise while the number of buyers increases proportionally in comparison with the number of sellers until equilibrium is restored (Ehrman, 2006:99).

### 3.2.3 Momentum

Momentum measures the rate at which prices change over a fixed period of time. The formula for momentum is given by Achelis (2001:196):

$$M = \frac{P_0}{P_t} \times 100, \quad (3.1)$$

where  $P_0$  is today's closing price and  $P_t$  is the closing price  $t$  days ago. In order to determine the momentum of a particular share price, the latest closing price is subtracted from the closing price  $t$  days ago. The result is a series of positive and negative values that are then plotted around a zero line (Meyers, 1994:151). If prices are falling and the momentum line is below zero and falling, the downtrend is gaining momentum. If prices are falling but the momentum line levels out and (or) begins to

rise, the downtrend is losing momentum. A bullish divergence would occur if the momentum line moved in the opposite direction of declining price activity.

The opposite is also true. In the case where prices are rising and the momentum line is above zero and rising, the uptrend is gaining momentum. When prices are rising but the momentum line levels out and (or) begins to fall, the uptrend is losing momentum (Murphy, 1999:230). Accordingly, a bearish divergence would occur if the momentum line moves in the opposite direction of rising price activity. Both bearish and bullish divergences at overbought/oversold levels are of considerable importance as they indicate a potential trend reversal.

### 3.2.4 The RSI

The RSI was developed by J. Welles Wilder Jr., and designed to address three flaws often associated with oscillators (Meyers, 1994:157). Oscillators have a tendency to move erratically at times. This occurs as old data points are no longer considered in the calculation of the oscillator. As the oscillator moves on to incorporate new data points, older data is no longer taken into consideration, resulting in a misleadingly high or low reading (Murphy, 1986:296). To address this limitation, the RSI employs a smoothing of days on which prices closed up and down. The vertical scale of oscillators is another flaw that is addressed by the RSI. Because the RSI moves only between 0 and 100, overbought and oversold conditions are easily identifiable (Achelis, 2001:297).<sup>16</sup> The third oscillator flaw that the RSI addresses, is the large amounts of data that need to be stored (Meyers, 1994:161).

The term 'relative strength' in this context, does not refer to the strength of a specific position relative to that of another position or index. Instead, it refers to the relative strength of the position relative to itself (Whistler, 2004:36). The RSI formula is given by Rogers (1999:98):

$$RSI = 100 - \frac{100}{1+RS}, \quad (3.2)$$

---

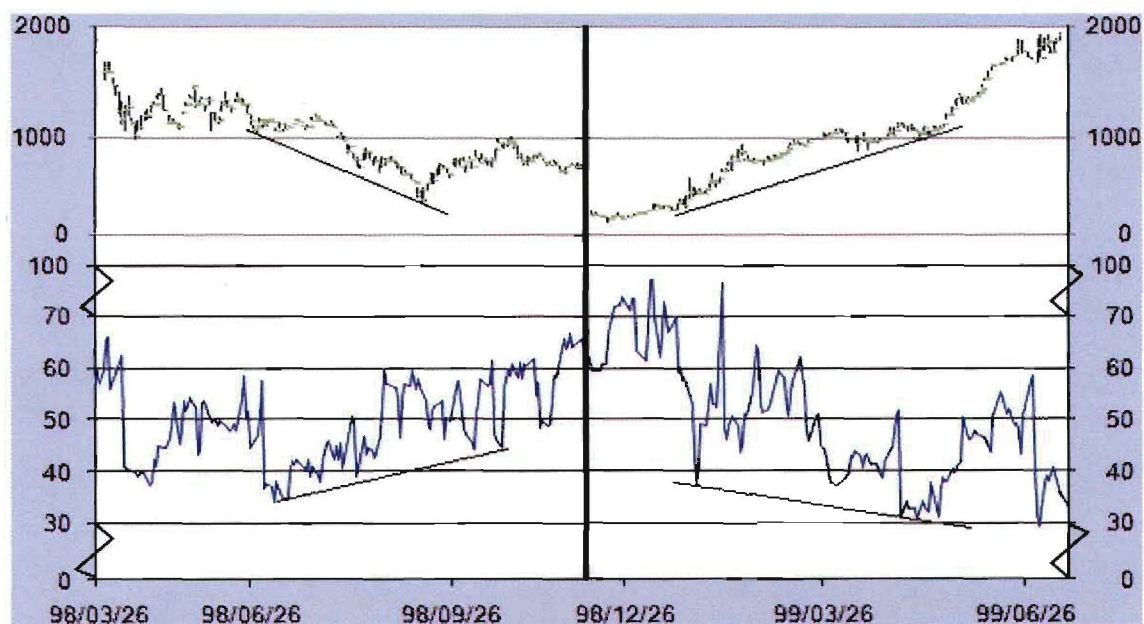
<sup>16</sup> The green line in figure 3.1 shows the actual values of the commodity while the blue line indicates the RSI values at each date.

$$\text{where } RS = \frac{\text{Average of net days' up closing prices}}{\text{Average of net days' down closing prices}}$$

The RSI is therefore the ratio of the average value of days with rising prices compared to the average value of days with falling prices. These values are all recalculated as a percentage from 0 percent to 100 percent. Values above the 70 percent mark indicate an overbought market, while values below the 30 percent boundary indicate an oversold market (Murphy, 1999:242). While a falling RSI value indicates that the ratio of average rising prices to falling prices is decreasing, a rising RSI value indicates that the ratio of average rising prices to falling prices is increasing. Accordingly, the 30 percent line is viewed as a trigger point for bullish signals, and the 70 percent line as a trigger point for bearish signals.

Figure 3.1 illustrates a hypothetical bullish divergence on the left-hand side and a bearish divergence on the right-hand side. A bullish divergence occurs when the RSI indicator moves in the opposite direction of declining price activity, while a bearish divergence occurs when the RSI indicator moves in the opposite direction of rising price activity (Meyers, 1994:150).

FIGURE 3.1 RSI BULLISH AND BEARISH DIVERGENCES



It should be noted that the 30 percent and 70 percent lines are not necessarily fixed boundaries regarding overbought and oversold conditions. Some traders widen the gap between the two boundaries to 20 percent and 80 percent (Ehrman, 2006:102). In addition to the changes made to the signal boundaries, the trader could also use a variety of timeframes to calculate the RSI. Most traders do, however, make use of the standardised timeframes of five, nine or fourteen days. The sensitivity of the oscillator increases when shorter timeframes are used (Murphy, 1986:298).

### 3.2.5 The Chande Momentum Oscillator

The Chande Momentum Oscillator (CMO) was introduced by Tushar Chande and Stanley Kroll in 1994. Although the CMO is closely related to the RSI, it differs from the RSI in that it is an unsmoothed momentum oscillator (Etzkorn, 1997:77).

The CMO is calculated by making use of unsmoothed data, which allows for the inclusion of extreme short-term movements. This attribute allows the oscillator to reach overbought and oversold positions more often (Pring, 2002:221). In deciding when an instrument is overbought or oversold, the CMO operates on a similar principal in that overbought and oversold positions are identified when the oscillator crosses certain levels or scales. However, the difference between the RSI and the CMO lies in the fact the CMO does not oscillate between 0 and 100, but between - 100 and + 100 (Colby, 2002:146). This means that the midpoint is not 50 but 0, which allows the trader to identify more clearly when momentum is positive and when it is negative.

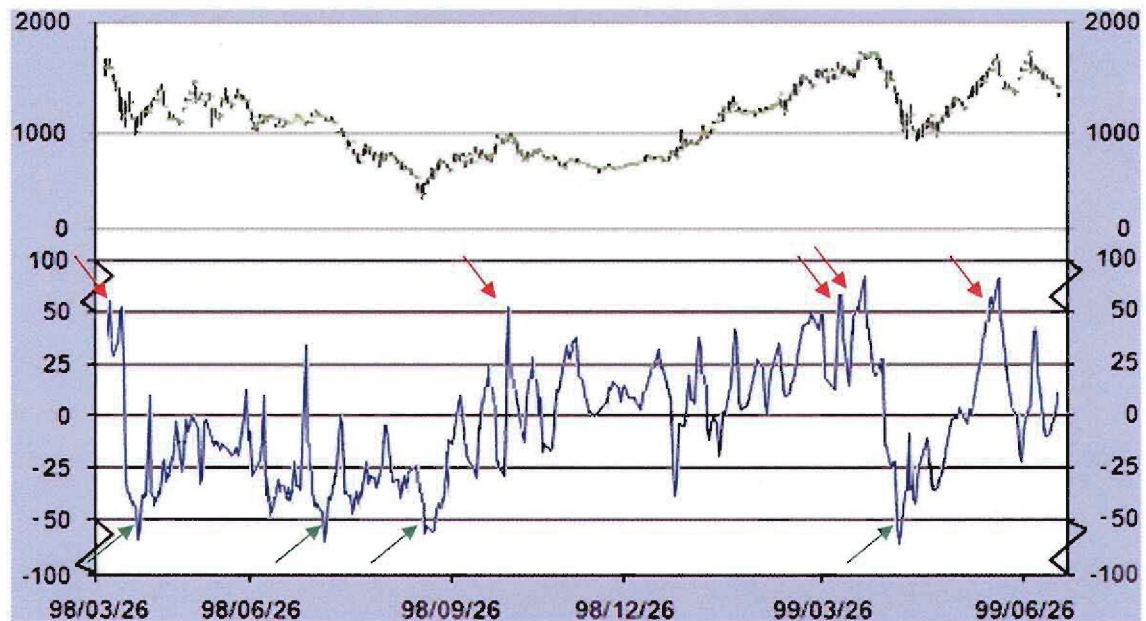
In order to calculate the CMO, the sum of up-days and down-days and the difference between up-days and down-days has to be calculated first. The calculation of the CMO can be given by:

$$CMO = 100 \times \left[ \frac{(S_u - S_d)}{(S_u + S_d)} \right], \quad (3.3)$$

where  $S_u$  is the sum of the (close-to-close) up-day momentum for  $n$  days, and  $S_d$  the sum of the (close-to-close) down-day momentum for  $n$  days.

As can be seen from the above calculation of CMO, it also differs from the RSI in another respect, in that it includes both up and down momentum in the numerator, where the RSI only includes upward momentum in the numerator. Figure 3.2 below illustrates how entry and exit points are identified by making use of the CMO. The red arrows indicates sell signals, while the green arrows indicate buy signals.

FIGURE 3.2 IDENTIFYING CMO ENTRY AND EXIT POSITIONS



Apart from its use as explained above, CMO can also be utilised for identifying trends. A higher absolute value of CMO indicates a stronger trend, while lower absolute values of the CMO indicate that the instrument is trading in a sideways trading range. Unfortunately, the CMO was not available on the software package used for identifying entry and exit points at the time of this study, and no results will be reported for this indicator.

### 3.2.6 The Stochastic Oscillator

The Stochastic oscillator, another market strength indicator used by technical analysts, was popularised by George Lane (Meyers, 1994:191). The Stochastic oscillator compares the level at which a commodity's price has closed relative to its price range over a specific time period (Weissman, 2005:32). This market strength indicator is based on the observation that closing prices tend to be closer to the upper end of the

price range amidst increasing prices. Conversely, closing prices tend to be near the lower end of the range during a downtrend (Murphy, 1986:304). As the trend matures and nears reversal in an upward trending market, the closing prices tend to be further from the high observed early in the trend. Conversely, the closing prices will move farther from the low earlier observed during the maturity and reversal level in a downward trending market (Ehrman, 2006:103).

The Stochastic oscillator aims to establish when prices begin to close at an increasing distance from their trading range highs. Based on this logic, a reversal in an upward trending market is imminent when the closing prices begin to group near the day's low. Similarly, when the closing prices are grouped near the day's high in a downward trending market, a trader can begin to look for a reversal in the trend (Ehrman, 2006:103).

The Stochastic oscillator is plotted as two lines, the %K line and the %D line (Achelis, 2001:321). The equation for the K line, which is the more sensitive of the two, is given by Murphy (1999:240):

$$\%K = 100 \left[ \frac{(C - L_t)}{(H_t - L_t)} \right], \quad (3.4)$$

where  $C$  is the latest closing price,  $L_t$  is the lowest low for the last  $t$  trading periods (days, weeks or months), and  $H_t$  is the highest high over the same period. The %D line is a simple moving average of the %K line (Rogers, 1999:100), as it is a smoother, shorter version of the K line, given by (Ehrman, 2006:103):

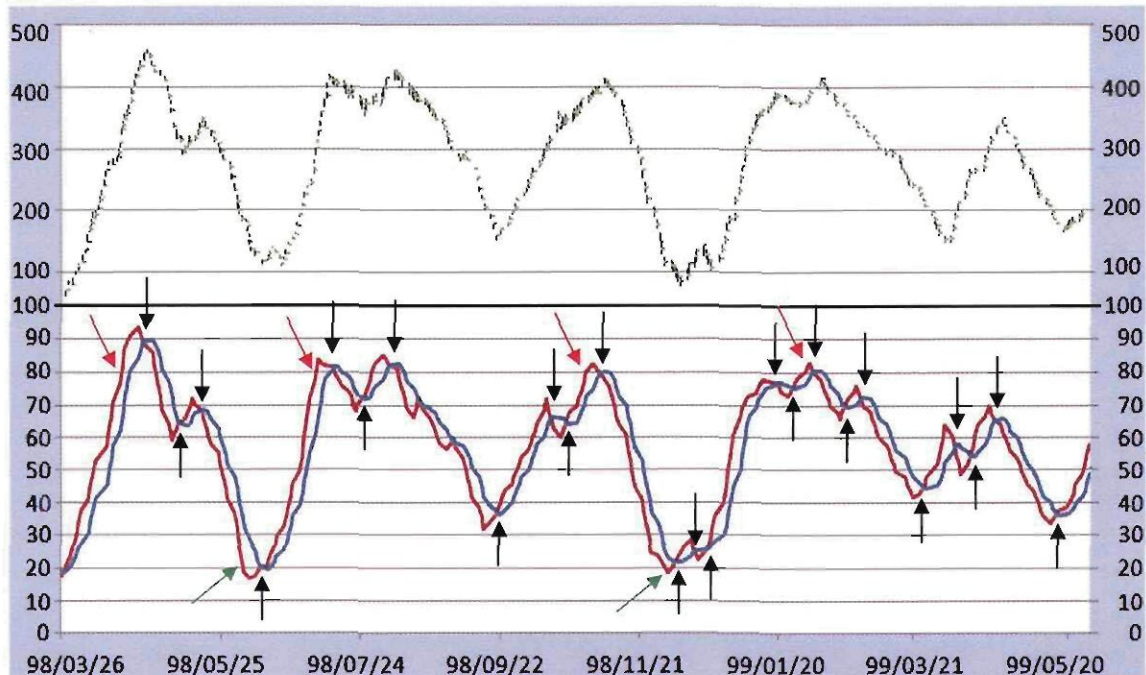
$$\%D = 100 \left[ \frac{H_t}{L_t} \right], \quad (3.5)$$

where  $H_t$  is the sum of  $(C - L_t)$ , and  $L_t$  is the sum of  $(H_t - L_t)$ . Since the %D line produces the major signals, it is regarded as the more important of the two lines.

As with the RSI, almost any number of days can be used for the testing of a trend over a given time period. This is because the intention is to establish where the most recent

closing price is in relation to the price range for a chosen period of time. Figure 3.3 illustrates a hypothetical Stochastic oscillator as adapted from Murphy (1999:256).

FIGURE 3.3 THE STOCHASTIC OSCILLATOR



In figure 3.3, the green line indicates the actual commodity prices, while the %K line is given in red and the %D line is given in blue. The green arrows in figure 3.3 indicate buying signals below the 20 percent line, while the red arrows indicate selling signals above the 80 percent line. The black arrows indicate buying and selling signals where the %K and %D lines cross.

The Stochastic oscillator can be employed in two ways to signal buying and selling opportunities. In a similar manner to the RSI, markets with readings above 70 – 80 percent are considered overbought, while markets with readings below 20 – 30 percent are considered oversold (Weissman, 2005:32). Alternatively, the Stochastic oscillator can be utilised to identify selling signals when the %D line crosses the %K line from below. Contrariwise, buying signals are identified when the %D line crosses the %K line from above.

A slowed version of the Stochastic oscillator is often employed in order to provide fewer but more reliable, signals. This slowed version is produced by taking another three-period average of the %D line (Ehrman, 2006:103). Because some traders prefer to trade over short periods of time, and others over longer periods, it is important that the trader decides on the time frame over which overbought and oversold conditions will be measured.

### 3.2.7 Conclusion

Both the Stochastic oscillator and the RSI are important tools used by traders in determining the relative degree to which a commodity or market is overbought or oversold. These tools can also be used to measure the overbought or oversold status between two different commodities. These market strength indicators are thus useful in the determination of entry and exit points, irrespective of the time frame employed. Both these indicators will be used in chapter 7. In the next section, moving average indicators are discussed in order to further the understanding of the pairs trading processes in chapter 7.

## 3.3 Moving Average Indicators

### 3.3.1 Introduction

Moving average indicators are widely used by technical traders (Achelis, 2001:27). A moving average can simply be constructed by finding the average price of a given commodity over a set period of time, and allowing this time frame to shift on as new data points are added. This information is then used by technical traders to determine how a share is behaving relative to its past behaviour (Ehrman, 2006:106). The two moving average indicators discussed for the purpose of pairs trading are Bollinger bands and Moving Average Convergence/Divergence (MACD). These indicators are not used alone, but in conjunction with the market strength indicators discussed in section 3.2.

Prior to discussing the different moving average indicators, the following section examines volatility and standard deviation as a statistical measures used in moving average indicators.

## 3.3.2 Volatility and Standard Deviation

### 3.3.2.1 Introduction

The distribution of the rates of return around an average rate of return is a popular measure of risk (Levy, 2002:134). In the case where the rate of return deviates far from the average, it means that the dispersion, or standard deviation, is large and there is relatively large volatility (risk) associated with such an investment (Levy, 2002:133). Even though standard deviation is a statistical measure of volatility, only volatility will be discussed here (Achelis, 2001:308). More will be said about standard deviation in section 3.3.3.

Volatility can be measured by means of a Simple Moving Average (SMA), Exponentially Weighted Moving Average (EWMA), and a GARCH model (Sin & Lam, 2004:16). Since GARCH will be explained in chapter 4, only SMA and EWMA are discussed here.<sup>17</sup>

SMA and EWMA are important to the pairs trader for two reasons. Firstly, SMA forms part of the Bollinger Band calculations. Secondly, both SMA and EWMA can be used as measures of volatility. Although no explicit reference will be made to volatility in chapter 7, these measures will be used in monitoring the pairs portfolio in chapter 7.

### 3.3.2.2 SMA

A SMA is merely the average of a set of variables (such as commodity prices) within a specific time period. The average is moving, thus discarding the oldest data point whilst simultaneously inserting the new data point into the calculation. By doing this, the data is smoothed from any spikes (Gitam Institute of Foreign Trade (GIFT), 2005:3). A simple moving average can be calculated as follows (Achelis, 2001:72):

$$SMA = \frac{\sum_{i=1}^n price_i}{n}, \quad (3.6)$$

---

<sup>17</sup> Please refer to chapter 4 and 5 for an in depth discussion on GARCH, and its use in identifying seasonal patterns in financial time series data.

where  $n$  is the number of time periods. In order to determine the  $n$ -period SMA volatility (or standard deviation) the  $n$ -period average of price returns must be determined first. After this has been done, the sum of the squares of the differences between each period's return, and that of the average is determined and measured over the full  $n$ -period. The quotient of this sum and  $n - 1$  is calculated, and the overall square root determined.

This model is the most widely used volatility model in Value at Risk (VAR) studies and is given by KPMG (1997:153):

$$h_t = \sqrt{\frac{1}{k-1} \sum_{s=n-k}^{n-1} (x_s - \mu)^2}, \quad (3.7)$$

where  $h_t$  is the Simple Moving Average volatility measure,  $k$  the number of observations,  $x_s$  the observation at time index  $n$ , and  $\mu$  the expected value of all the observations.

One drawback of SMA is that although this method of determining risk accounts for the increase or decrease in the price, it does not manifest itself quantitatively in the SMA. The reason for this being that the time order of the data (observations) is not accounted for when using the SMA calculation to calculate volatility.

A further drawback of SMA is that an equally weighted average is used. Since recent data is far more important for volatility forecasting than older data points, this method of volatility measurement is not adequate. It is, however, still relevant to the Bollinger Band calculations discussed in section 3.3.3.

### 3.3.2.3 EWMA

To address the problem of equally weighted data points, an EWMA as a measurement tool for volatility can be used. By making use of the EWMA method, the more recent data is given a larger weighting whilst older data points receive a smaller weighting. The EWMA model assigns the most weight to the previous day's value, which in turn is

dependent on the day before, and so on (Pafka & Kondor, 2001:3, also Goodworth & Jones, 2004:5).

This model is thus superior to the SMA model in that it is able to capture the more recent and therefore, more important data for calculating volatility. A value closer to 0 indicates that the most recent data has a larger impact on current volatility, while a value closer to 1 indicates that the most recent observation has a smaller impact on current volatility.

The formula of the EWMA model is given by (KPMG Risk, 1997:153):

$$h_t = \sqrt{(1 - \lambda) \sum_{s=t-k}^{t-1} \lambda^{t-s-1} (x_s - \mu)^2}, \quad (3.8)$$

where  $h_t$  is the  $t^{\text{th}}$  period volatility,  $\lambda$  is the decay factor,  $(x_s - \mu)^2$  is the  $(t - 1)^{\text{th}}$  period squared return.

While EWMA is considered to be a superior model if compared with SMA models, it might happen that data is heteroskedastic. Data, in which the error terms may reasonably be expected to be larger for some points or ranges of the data than for others, are considered heteroskedastic (Engle, 2001:160). Although the regression coefficients for an ordinary least squares regression is still unbiased when working with heteroskedastic data, the standard errors and confidence intervals estimated by conventional procedures will be too narrow, thus providing a false sense of precision. ARCH and GARCH models can be used to address this problem.

### 3.3.2.4 Conclusion

Even though SMA and EWMA are often used to indicate the level of volatility associated with a particular commodity, or portfolio, these measures might not be the best to use. Since SMA is part of the Bollinger Band calculation it is relevant for the pairs trader. EWMA exceeds SMA as a measure of volatility, but it too will not provide a true reflection of the volatility associated with a particular commodity or portfolio.

This is specifically the case if the data is heteroskedastic. The GARCH models discussed below address the problem of heteroskedasticity.

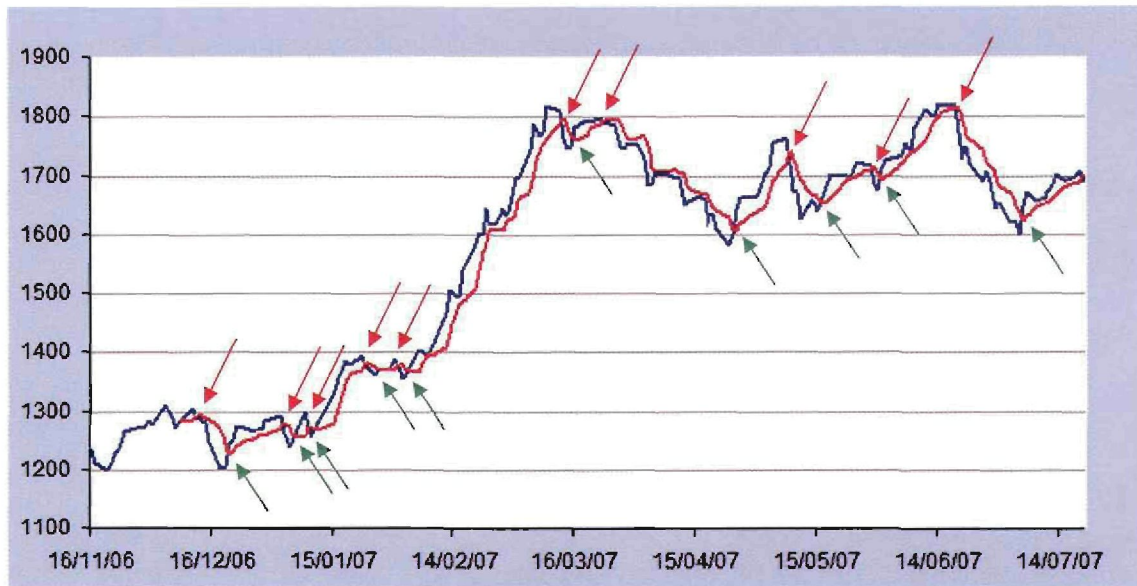
### 3.3.3 Moving Averages as Indicator

Even though moving averages (MAs) are mostly used in a variety of other technical indicators, MAs can also be used on their own to convey important information regarding the market. In very volatile markets, MAs are usually used as indication of the trend by smoothing the data (Masonson, 2004:133). When utilising MAs in this manner, buy signals are usually identified when the MA Moving averages can also be used in conjunction with other moving averages (moving averages of different time frames) to indicate entry and exit points. Another way of using moving averages to indicate entry and exit points, is to use only one moving average of the price series in conjunction with the price series from which it was derived.

When making use of two moving averages of different time frames (for example a 5 day MA and a 20 day MA), entry and exit points are usually identified when these moving averages cross. If the 5-day MA cross the 20-day MA from beneath, a buy signal is registered. In the case of the 5-day MA cross the 20-day MA from above, a sell signal is identified.

In making use of moving averages in conjunction with the original price series, buy and sell signals are identified in a similar manner. A buy signal will typically be identified when the original price crosses the moving average from below, while a sell signal will be identified when the original price series cross the moving average from above. For the purpose of this thesis, MAs will only be used in conjunction with the original price series to identify buy and sell signals for the commodity trader. Figure 3.4 below illustrates how MAs will be used for the purpose of this thesis.

FIGURE 3.4 MOVING AVERAGES



As before, the red arrows indicate sell signals, while the green arrows indicate buy signals. Sell signals are generated when the blue line (original price series) cross the red line (the MA) from above. Conversely, buy signals are generated when the price series brakes through the MA from beneath.

### 3.3.4 Bollinger Bands

Bollinger bands are two trading bands that surround the moving average of a particular share's price. The Bollinger bands are plotted two standard deviations away on either side of the simple moving price average (usually a twenty-day SMA) (Murphy, 1999:209). As mentioned in section 3.3.2.1, standard deviation is considered a measure of volatility. Standard deviation describes how prices are dispersed around an average value and is given by (Thomas, 2005:12):

$$\hat{\sigma} = \sqrt{\frac{\sum x_i^2}{n}}, \quad (3.9)$$

where  $\hat{\sigma}$  is the standard deviation at the  $i^{\text{th}}$  observation.

For the purpose of using standard deviation, as defined in section 3.3.2.1, for Bollinger bands, equation 3.8 is manipulated to take the form of (Achelis, 2001:309):

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (Close_i - n\text{-period SMA of Close})^2}{n}}, \quad (3.10)$$

where SMA is the Simple Moving Average (see equation 3.5 in section 3.3.2.2 above), and  $n$  is the number of time periods. Since the standard deviation is a measure of volatility, the Bollinger bands adjust themselves to the current volatility in the market. When market conditions are more volatile the Bollinger bands will widen, and when market conditions are less volatile, they will contract (Achelis, 2001:71). The equations for the upper and the lower bands are given by Achelis (2001:74):

$$Upper\ Band = Middle\ Band + \left[ D \times \sqrt{\frac{\sum_{i=1}^n (Close_i - Middle\ Band)^2}{n}} \right], \quad (3.11)$$

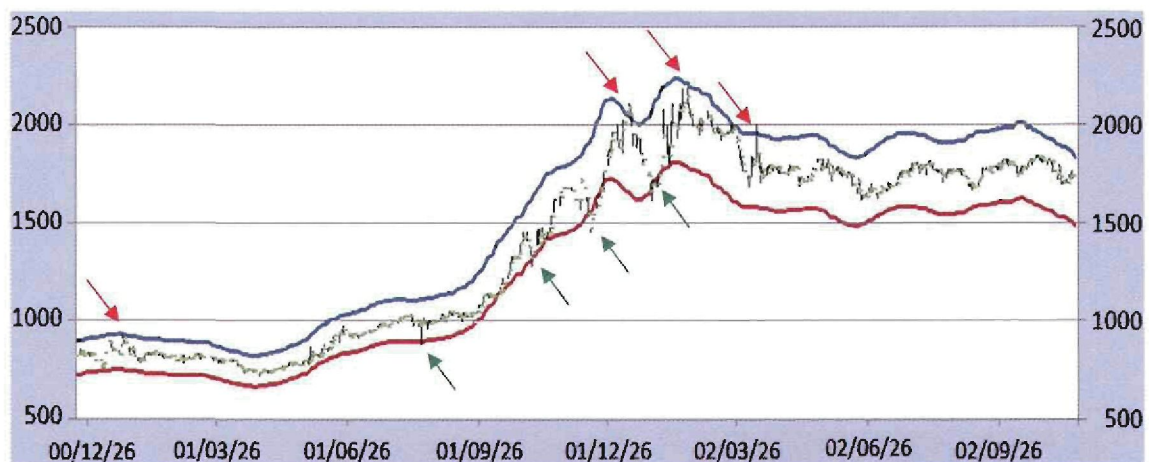
$$Lower\ Band = Middle\ Band - \left[ D \times \sqrt{\frac{\sum_{i=1}^n (Close_i - Middle\ Band)^2}{n}} \right], \quad (3.12)$$

where  $D$  is the standard deviation, and the middle band is equal to the SMA over the time period. The upper band is the same as the middle band, but it is shifted up by the number of standard deviations. The lower band is similar to the middle band, but is shifted down by the number of standard deviations. By using two standard deviations, 95 percent of the price data will fall between the two trading bands. As is the case with the RSI, prices are considered overbought when they touch the upper band, and oversold when they touch the lower band (Murphy, 1999:209).

When using longer time periods, many technical traders increase the distance from the moving average to 2.5 standard deviations. Technical traders may, however, decrease the distance between the trading bands and the moving average from 2 standard

deviations to 1.5 standard deviations when trading over a shorter time period.<sup>18</sup> Figure 3.5, adapted from Murphy (1999:209), illustrates hypothetical Bollinger bands drawn around an invented price series depicted on the y-axis. The green line in figure 3.5 depicts the actual commodity prices, while the blue line indicates the upper band, and the red line indicates the lower band. The green arrows indicates buy signals, while the red arrows indicates sell signals.

FIGURE 3.5 BOLLINGER BANDS



When interpreting Bollinger bands, it is easier to use the upper and lower bands as price targets. If the price moves up from the lower band to cross the moving average,<sup>19</sup> the upper band becomes the upper price target. Conversely, a crossing below the moving average signifies the lower band as the lower price target. During a strong uptrend, prices will usually fluctuate between the upper band and the moving average. A crossing below the moving average will thus signal a trend reversal (Murphy, 1999:210).

The band width of Bollinger bands can also be used as an indication of possible trend reversals. If the bands stay unusually narrow (around the moving average), it is usually a sign that a new trend is about to start. Once the bands then widen (caused by increased

<sup>18</sup> The standard time period for Bollinger bands is 20 periods (days, weeks, months). When trading over a longer time period, traders make use of a 50-day (week or month) moving average, and, when traders trade over a shorter time period, they make use of a 10-day (week or month) moving average (Ehrman, 2006:107).

<sup>19</sup> Please note that this moving average (middle band) is not included in this illustration.

volatility), and stay unusually far apart, it is often a sign that the current trend is ending (Murphy, 1999:211).

Finally, it is to be noted that Bollinger bands are best used when combined with overbought and oversold oscillators (GIFT, 2005:3). Bollinger bands, as with most other statistical tools, only reach their full predictive potential when they are used in conjunction with other statistical tools (Ehrman, 2006:107).

### 3.3.5 MACD

The MACD indicator, developed by Gerald Appel, is one of the simplest and most trustworthy indicators available (GIFT, 2005:4). The MACD indicator is a trend-following momentum indicator that shows the relationship between two moving averages of prices. The first line, which is called the MACD line, moves faster than the second, which is called the signal line. The MACD line is the difference between the exponentially smoothed moving averages of closing prices over a given period.<sup>20</sup> The signal line is the exponentially smoothed average of the trailing nine MACD periods (days, weeks or months) (Murphy, 1986:312). The MACD indicator thus combines some principles of oscillators with a dual moving average crossover approach.

There are three popular applications of the MACD line: overbought/oversold conditions, crossovers and divergences (Achelis, 2001:199). Where the MACD line is used to indicate overbought/oversold conditions, a signal is generated when the shorter moving average pulls away dramatically from the longer moving average, thus indicating that the price is overextended, and will soon move to more market related levels. If this occurs, a buy signal is generated when the MACD line crosses the signal line from below when both have negative values, that is both lines are below the zero line. Conversely, a sell signal is generated when the MACD line crosses the signal line from above when both have positive values, that is both lines are above the zero line (Rogers, 1999:104).

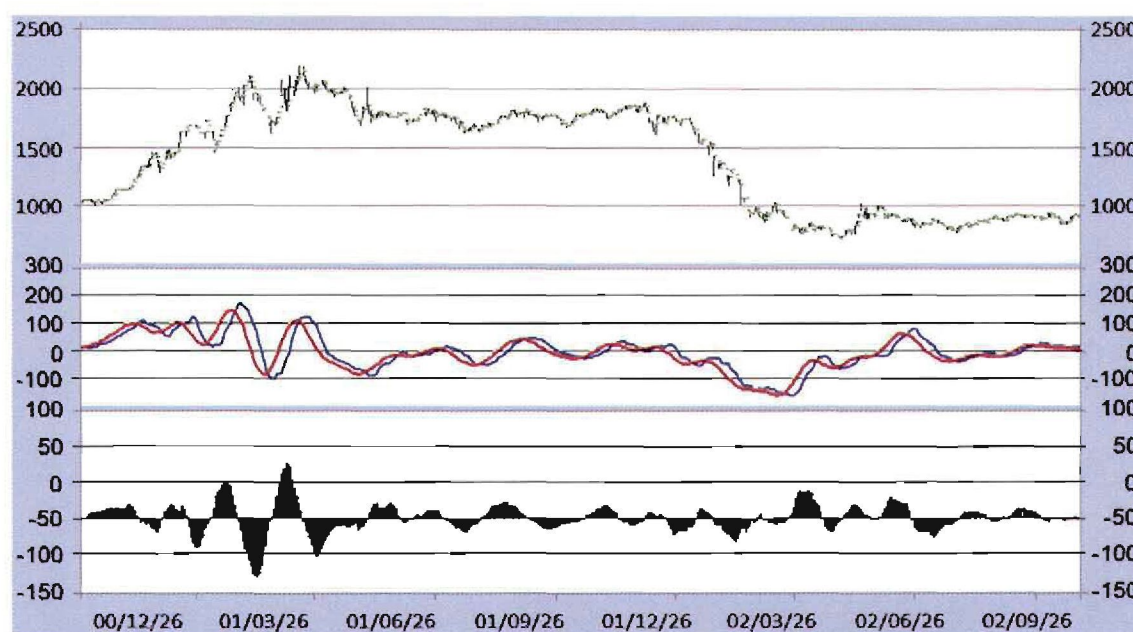
---

<sup>20</sup> The 12-day and 26-day moving averages are the most popular (Ehrman, 2006:108). The fast MACD line is generated by subtracting the 26-period Exponentially Moving Average (EMA) from that of the 12-period EMA.

When using the MACD line for crossover purposes, buy signals are generated when the MACD line rises above the signal line and sell signals are generated when the MACD line falls below the signal line (Achelis, 2001:200). Buy and sell signals are similarly also generated when the MACD line moves above or below zero. Because the magnitude of the price movement that follows the trend reversal tends to be proportionate to the distance that the MACD line is from zero, the extent of the price move can be derived from the MACD line. Consequently, it follows that the movement of prices after a trend reversal tends to be greater in magnitude where the MACD line is far below or far above zero (Ehrman, 2006:109).

If the MACD is used to indicate a trend reversal, the trader focuses on the divergence of the MACD line from the commodity price. When the MACD line achieves new highs while the price of the commodity is falling, a bullish divergence is present. If the MACD line reaches new lows while the price of the commodity is rising, a bearish divergence is present in the market (Murphy, 1999:254).<sup>21</sup> Figure 3.6, below illustrates hypothetical MACD lines derived from an invented price series. Note that the green line indicated on the y-axis in figure 3.6 shows the actual commodity prices. In the centre graph, the MACD line is indicated by the red line, while the signal line is indicated by the blue line. The bottom graph in figure 3.6 shows the MACD in histogram form.

FIGURE 3.6 THE MACD INDICATOR



<sup>21</sup> See section 3.2.4 for an explanation of bullish and bearish divergences.

The MACD indicator can be improved in the form of a histogram to give earlier warning signals than the MACD lines. Since buy and sell signals are generated by the crossing of the MACD line and signal lines, the difference between the lines will provide a single line that oscillates around zero. The histogram value will be positive (histogram values above the zero line) when the MACD line crosses the signal line from below. In this case, the two lines are in a bullish alignment. Conversely, the histogram value will be negative (histogram values below the zero line) when the MACD line crosses the signal line from above. In this case, the two lines are in a bearish alignment (Murphy, 1996:126).

### 3.3.6 Conclusion

Moving average indicators are fairly versatile, and are often used for more than one purpose. Of the many moving average indicators, Bollinger bands and MACD are two of the most widely used indicators. Bollinger bands can be used to identify the direction of a trend, recognise trend reversals, and to indicate the volatility of a price series. MACD, on the other hand, is often used to indicate overbought and oversold conditions, crossovers, and divergences. Even though these indicators are useful on their own, they are more effective when used in conjunction with other technical tools.

## 3.4 Summary

Even though it is not possible to trade successfully without bearing the fundamentals of the agricultural commodity market in mind, it is imperative that the pairs trader have a thorough understanding of technical analysis. This knowledge is essential in deciding when positions should be opened and closed, when managing a pairs portfolio.

However, there are a multitude technical tools available, and it is easy to get lost in the process of deciding on the best indicators. Because it is impossible, and impractical, to make use of all the technical analysis tools available, the trader will do well to choose only a few of these indicators with which he is comfortable. For the purpose of this thesis, attention was only paid to the most widely used indicators: the RSI, The Stochastic oscillator, Bollinger bands and MACD.

These technical analysis tools will be used (chapter 7) in conjunction with the fundamental analysis discussed in chapter 2. In addition to these two analyses, tests will also be performed to establish whether seasonal patterns can be identified from the data. Chapter 4 below will begin with a discussion on the statistical measures that will be used to test for seasonality in the agricultural commodities data. Three methods of seasonality tests will be discussed. The first two, ARIMA and GARCH, are used to identify observed seasonal patterns, while UCM modelling will be employed to identify unobserved seasonality.

# CHAPTER 4

## Statistical Measures used for Testing Seasonality

*An investor armed with good information and analysis can often move ahead of the crowd.*

(McHattie, 2007:46)

### 4.1 Introduction

Chapter 2 has explored the history of several agricultural commodities currently being traded on SAFEX, as well as some of the factors that influence the prices of these commodities. In order to assemble a profitable pairs portfolio, the trader must be aware of seasonal patterns in the data. It is, therefore, imperative that the fundamental analysis of such a pairs portfolio is linked to an empirical analysis to determine what seasonal patterns exist, and how these patterns may be utilised in a pairs trading strategy.

Chapter 4 elucidates the statistical procedures that are used in chapter 5, regarding the testing for seasonal patterns in the agricultural commodities market. These seasonal patterns will be instrumental during the pairs selection process where the statistical relationship between seasonal patterns are assessed to determine whether two series can be traded successfully as pairs.

In order to test for seasonality in the data, several statistical models need explaining. The chapter commences with section 4.2, in which various studies that have been conducted on seasonality in economic data are reviewed. Thereafter, the most prominent and important statistical models used in these studies for testing seasonality are introduced for further discussion: the ARMA model and the ARIMA model in section 4.3, the GARCH model in section 4.4, and UCM modelling in section 4.5.

## 4.2 Seasonality

### 4.2.1 Introduction

Seasonality is best described as different average returns for all periods, distinguishable as a distinct seasonal component, cyclical component, trend, and irregular component (Lee, Lee & Lee, 2000:798). The literature distinguishes between various seasonal effects, including a 'month-of-the-year effect', a 'day-of-the-week effect' and a 'weekend effect' (Berument & Kiyamaz, 2001:1). The day-of-the-week effect was first observed and documented by M.F.M. Osborne, the physicist who applied the concept of Brownian motion to the stock market in 1959 (Osborne, 1959).

### 4.2.2 Literature Review

Since Osborne's discovery in 1959, Cross (1973), French (1980), Gibbons and Hess (1981), Lakonishok and Levi (1982), Keim and Stambaugh (1984), and Rogalski (1984), among others, confirm that there are differences in the distributions of stock returns in each of the weekdays. The results of these studies indicate that the average return on Mondays is considerably less than the average return during the other weekdays.

Although these studies were performed on the equity markets in the United States, the day-of-the-week effect has also been investigated for both international equity markets and international non-equity financial markets. Jaffe and Westerfield (1985a, 1985b) find significant negative mean returns on Mondays in the United States, Canada and the United Kingdom stock markets, and significant negative Tuesday returns in the Japanese and Australian stock markets. Aggarwal and Rivoli (1989) study the emerging markets of East Asia and observe lower mean returns on Mondays and Tuesdays in the stock returns of Hong Kong, Singapore, Malaysia and the Philippines, from September 1976 to June 1988.

Kato and Schallheim (1985); Chang, Pinegar and Ravichandran (1993); Athanassakos and Robinson (1994); and Dubois (1986) show that the distributions of stock returns also vary by the weekdays internationally. The day-of-the-week effect is also detected

in the commodity and stock futures markets (Cornell, 1985; Dyl & Maberly, 1986; Gay & Kim, 1987), the Treasury Bill market (Flannery & Protopapadakis, 1988), and in the foreign exchange market (Corhay, Fatemi & Rad, 1995).

While the focus of the above studies has been the seasonal pattern in mean returns, there are several other empirical studies investigating the time series behaviour of stock prices in terms of volatility by using variations of GARCH-type models (French, Schwert & Stambaugh, 1987; Akgiray, 1989; Baillie & DeGennaro, 1990; Hamao, Masulis & Ng 1990; Nelson, 1991; Campbell & Hentschel, 1992; and Ogum, Nouyrigat & Beer 2002). These studies report that the expected returns in stock markets are time-varying and conditionally heteroskedastic. By making use of several GARCH models, Heymans (2005) also finds a day-of-the-week effect in the South African stock index returns.

In terms of monthly effects, a number of studies find year-end effects in equity markets (Keim, 1983; Reinganum, 1983). The existence of a seasonal effect is also found in the Australian stock market (Brown, Keim, Keleidon & Marsh, 1983), the United Kingdom stock market (Lewis, 1989), the Canadian stock market (Berges, McConnell & Schlarbaum, 1984), the Indian stock market, and Malaysian stock market (Pandey & Chotigeat, 2005).

Although in the minority, work has also been undertaken that makes use of structural time series analysis (STA) for identifying unobserved or hidden seasonal patterns. The UCM modelling analyses forecasts equally-spaced univariate time series data, using Unobserved Components Models.

The UCMs, also called 'Structural Models' in the time series literature, divide the response series into four components. These components are meant to capture the salient features of the series that are useful in explaining and predicting its behaviour (SAS, 2006a:1619). The four components are comprised of a trend, a seasonal component, a cyclical component, and the regression effects due to predictor series. Harvey (1990) terms the components in a UCM the 'stylised facts' about the series under consideration.

UCM modelling has gained popularity in recent years. After Harvey and Scott (1994), made use of STA to test United Kingdom consumption growth, Al-Saad and Moosa (2005) investigated the nature of seasonality in the monthly stock returns derived from a general index of the Kuwait Stock Exchange. Al-Deehani (2006) furthered this by investigating seasonality factors for the returns of the Kuwait stock market and its various sectors.

### 4.2.3 Conclusion

From the brief literature review above, it is clear that many tests have been performed in an attempt to identify seasonal patterns in financial time series data. For the purpose of this thesis, only three types of models are reviewed: the ARMA/ARIMA-type models, GARCH, and UCM.

Section 4.3 discusses the ARMA and ARIMA processes, and the use of these models in time series analysis. In addition to this, commentary is also provided on the use of these models in the study of seasonality in time series data. In section 4.4, GARCH is introduced and discussed with reference to its statistical characteristics. Thereafter, UCM is discussed in section 4.5.<sup>22</sup>

## 4.3 An Introduction to ARMA and ARIMA

The ARMA model is the result of a combination of the AR and MA processes. A mixed autoregressive and moving-average process containing  $p$  AR terms and  $q$  MA terms is said to be an ARMA process of the order  $(p,q)$ . The general ARMA( $p,q$ ) is given as (Tsay, 2002:50):

$$R_t = \varphi_0 + \sum_{i=1}^p \varphi_i r_{t-i} + \alpha_t - \sum_{i=1}^q \theta_i \alpha_{t-i}, \quad (4.1)$$

where  $\alpha_t$  is a white noise series and  $p$  and  $q$  are non-negative integers. By making use of the lag operator  $L$ , the above equation can be written as (Maddala, 2002:522):

---

<sup>22</sup> It is to be noted that these models are only introduced in this chapter. Each model will be discussed in more detail during the model selection process performed in chapter 5.

$$\phi(L)X_t = \theta(L)\varepsilon_t, \quad (4.2)$$

where  $\Phi(L)$  and  $\theta(L)$  are polynomials of orders  $p$  and  $q$ , respectively, and defined as:

$$\phi(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p \quad (4.3)$$

$$\theta(L) = 1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q. \quad (4.4)$$

The polynomial  $1 - \alpha_1 L - \dots - \alpha_p L^p$  is the AR polynomial of the model and  $1 + \beta_1 L + \dots + \beta_q L^q$  the MA polynomial (Maddala, 2002:523).

By allowing the AR polynomial,  $1 - \alpha_1 L - \dots - \alpha_p L^p$ , in the ARMA model to have 1 as a characteristic unit root, the model is transformed into an ARIMA model. The standard ARIMA model can be given by (Maddala, 2002:523):

$$W_t = \mu + \frac{\theta(L)}{\phi(L)} \varepsilon_t, \quad (4.5)$$

where  $W_t$  is the response series,  $\mu$  is the mean term,  $L$  is the backshift operator,  $\phi(L)$  is the AR operator given by  $1 - \alpha_1 L - \dots - \alpha_p L^p$ ,  $\theta(L)$  is the MA operator given by  $1 + \beta_1 L + \dots + \beta_q L^q$ , and  $\varepsilon_t$  is the error term. The order of an ARIMA model is usually denoted by the notation  $\text{ARIMA}(p, d, q)$ , where  $p$  is the order of the AR element,  $d$  is the order of differencing, and  $q$  is the order of the MA process.

The ARIMA model is said to be unit root non-stationary because its AR polynomial has a unit root (Tsay, 2002:59). These roots lie on the boundary of or inside the unit root circle. In order to remove unit root non-stationarity the data series can be differenced (Koop, 2000:124). A differenced series is simply the difference between two data points, or more formally  $y_t = x_t - x_{t-1}$  (Ramanathan, 2002:518). However, differencing a series may amplify the effects of measurement errors in the original data, which will affect the estimates of the autocorrelation and partial autocorrelation functions (Zeller & Palm, 2004:19).

Harvey (1997:193 & 194) is opposed to making use of the ARIMA approach for time series analysis for two main reasons. His first reason is that it is hard to select the appropriate model (the orders) in testing for seasonality. It is difficult to identify complex ARMA processes in small samples, and even more difficult to identify the theoretical autocorrelation function from the sample autocorrelation function. In the case of non-stationary data, (thus mostly differenced data) the selection of an inappropriate model will be problematic for forecasting.

Harvey's (1997) second reason is that even once a model has been selected it may be difficult to interpret the results. This is because many of the models have no sensible interpretation, and give forecast functions that may have undesirable properties. This compounds the first problem and as a result it is difficult to select the correct model and possible to choose a problematic model instead.

In order to avoid these difficulties, modelling techniques such as STA or GARCH could be used. ARIMA-type models will be used for more than one purpose in this thesis. The first use will be to test for seasonality in the agricultural commodities data, and the second use will be to derive seasonally adjusted series of the various commodity series. These seasonally adjusted series will be compared with each other during the pairs selection process. The subsequent section (4.4) briefly discusses GARCH, which is followed by a discussion on STA modelling in section 4.5.

## 4.4 An Introduction to GARCH

GARCH can be regarded as a non-standard ARMA model for the  $\alpha_t^2$  series (Tsay, 2002:48). The GARCH model is a specialised form of the ARCH model. Engle's (1982) ARCH model permits the variances of the forecasted return terms to change with the squared lag values of the previous error terms, treating heteroskedasticity as a variance to be modelled. As a result, a prediction is computed for the variance of each error term. This is called an ARCH ( $q$ ) and is expressed by Nelson (1991:348):

$$\sigma_t = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2. \quad (4.6)$$

One of Engle's graduate students, Tim Bollerslev, introduced a model that expanded the ARCH model soon after its publication by Engle (1982). Bollerslev called it the generalized ARCH or GARCH model and it takes the following basic form (Kungliga Vetenskapsakademien (KVA), 2003:15):<sup>23</sup>

$$\sigma^2_t = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2. \quad (4.7)$$

The first-order ( $p = q = 1$ ) GARCH model has since become the most popular ARCH model in use (KVA, 2003:15). The GARCH model essentially generalises the purely autoregressive ARCH model to an autoregressive moving average model. The weights on past squared residuals are assumed to decrease geometrically at a rate to be estimated from the data. Compared to Engle's basic ARCH model, the GARCH model is a helpful technical innovation that allows a parsimonious specification; a first-order GARCH model contains only three parameters.<sup>24</sup>

GARCH allows the user to model the serial dependence of volatility by including past variances in the explanation of future variances. The most widely used GARCH specification asserts that the best predictor of the variance in the next period is a weighted average of the long-run average variance, the variance predicted for this period, and the new information that is captured by the most recent squared residual in this period (Engle, 2001:160).

GARCH is thus considered a time-varying volatility measurement model that captures the effect of changing volatility over time. Pronounced temporal volatility clustering forms because financial return volatility data is influenced by time-dependent information flows. These time series data points can be parameterised using GARCH models. Although GARCH also considers excess kurtosis or "fat tail behaviour"

<sup>23</sup> It is to be noted that both the standard ARCH and GARCH models above have been adapted from the original texts in order to promote uniformity.

<sup>24</sup> This specification requires that  $\sum_{j=1}^q \alpha_j + \sum_{j=1}^p \beta_j < 1$ , in order to satisfy the non-explosiveness of the conditional variances. Furthermore, each of  $\alpha_0$ ,  $\alpha_j$  and  $\beta_j$  has to be positive in order to satisfy the non-negativity of conditional variances for each given time  $t$ .

(Hamao *et al.*, 1990:286), it is specifically designed to model time-varying conditional variances, and therefore often fails to capture highly irregular events.<sup>25</sup>

In his paper, Nelson (1991) recommends a different kind of GARCH model in order to eliminate some of the restrictive characteristics that make the common GARCH models unfit for explaining changes in volatile stock market returns. He (1991) finds that GARCH models may unduly restrict the dynamics of the conditional variance process because they impose parameter restrictions that are often violated by estimated coefficients. Moreover, it is difficult to interpret whether or not shocks to conditional variance persist, since the usual norms for measuring persistence often do not agree.

Nelson (1991) explains how changes to Bollerslev's GARCH model (equation 4.8 below) can enhance it to eliminate the abovementioned shortcomings. In the first phase, Nelson (1991:348) states that Bollerslev's GARCH model adds to Engle's original ARCH model by making the variance parameter ( $\sigma_t^2$ ) linear in lagged values of the error term  $\varepsilon_t^2$ , where  $\varepsilon_t^2 = \sigma_t^2 z_t^2$ . This model is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \beta_i \sigma_{t-i}^2 + \sum_{j=1}^p \alpha_j z_{t-j}^2 \sigma_{t-j}^2, \quad (4.8)$$

where  $\omega$ ,  $\alpha_j$  and  $\beta_i$  are non-negative. By adding a risk factor,  $\delta$ , the GARCH-M model of Engle and Bollerslev (1986a) is formed. This model takes the form of (Nelson 1991:348):

$$R_t = \varphi + \delta \sigma_t^2 + \varepsilon_t, \quad (4.9)$$

where  $\sigma_t^2$ , the conditional variance of  $R_t$ , enters the conditional mean of  $R_t$  as well. If for example,  $R_t$  is the return on a portfolio at time  $t$ , its required rate of return may be linear to its risk as measured by  $\sigma_t^2$ .

---

<sup>25</sup> The distributions of many stocks or indices have fat tails and therefore deviate from the normal distribution.

In the second phase, Nelson (1991) extends the GARCH(1,1) model to a multivariate context, and shows how Bollerslev, Engle and Wooldrige (1988) test a conditional capital asset pricing model (CAPM) with time-varying covariances of asset returns. By substituting recursively for the  $\beta_i \sigma_{t-i}^2$  terms, this model can assume the following form (Nelson 1991:348):

$$\sigma_t^2 = \omega^* + \sum_{k=1}^{\infty} \phi_k z_{t-k}^2 \sigma_{t-k}^2. \quad (4.10)$$

It follows that if  $\omega$ ,  $\alpha_j$  and  $\beta_i$  are non-negative,  $\omega^*$  and the  $\phi_k$  are also non-negative. By setting the conditional variance equal to a constant plus a weighted average (with positive weights) of past squared residuals, GARCH models capture the volatility clustering in asset returns. This characteristic makes GARCH models ideal for testing volatility patterns in market returns.

Nelson (1991) explains why this integrated model structure still imposes significant limitations on GARCH models. He states that there is evidence that stock returns are negatively correlated with changes in returns volatility, volatility tends to rise in response to 'bad' news and fall in response to 'good' news (Nelson 1991:349). He addresses this with his exponential GARCH (EGARCH) model.

It is further explained that the reason for these limitations is that GARCH models assume that only the magnitude and not the positivity or negativity of unexpected excess returns determines the feature  $\sigma_t^2$  (Nelson, 1991:349). If the distribution of  $z_t$  is symmetric, the change in variance tomorrow is conditionally uncorrelated with excess returns today. In equation 4.8 above,  $\sigma_t^2$  is a function of lagged  $z_t^2$ , and therefore invariant to changes in the algebraic sign of the  $z_t$ s. Therefore, it can be accepted that only the size and not the sign of lagged residuals determines conditional variance. This suggests that a model in which  $\sigma_t^2$  responds asymmetrically to positive and negative residuals might be preferable for asset pricing applications. This is significant for testing volatility patterns in agricultural commodity returns.

Nelson (1991) points out a second limitation of GARCH models resulting from the non-negativity constraints on  $\omega^*$  and the  $\phi_k$  in equation 4.10 above, which are imposed to ensure that  $\sigma_t^2$  remains non-negative for all  $t$ -values with probability 1. These constraints imply that increasing  $z_t^2$  in any period increases  $\sigma_{t+m}^2$  for all values of  $m \geq 1$ , ruling out random oscillatory behaviour in the  $\sigma_t^2$  process.

A third shortcoming of GARCH modelling concerns the interpretation of the ‘persistence’ of shocks to conditional variance. If volatility shocks persist indefinitely they may shift the whole term structure of risk premia, and therefore are likely to have a significant impact on investment in long-lived capital goods. It is found that in GARCH(1,1) models, shocks may persist in one norm and die in another; thus the conditional moments of GARCH(1,1) may explode, even when the process itself is strictly stationary (Nelson, 1990a). This may lead to difficulties when the parameters for the model are to be determined.

Nelson (1991) makes use of a different approach to ensure that  $\sigma_t^2$  remains non-negative. This must be done to ensure that these models capture the volatility patterns in market returns. To accomplish non-negativity,  $\ln(\sigma_t^2)$  is made linear in some function of time and lagged  $z_t$ 's for some suitable function  $g$ , given as (Nelson 1991:350) :

$$\ln(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}), \quad \beta_1 \equiv 1, \quad (4.11)$$

where  $\{\alpha_t\}_{t=-\infty, \infty}$  and  $\{\beta_k\}_{k=1, \infty}$  are real, non-stochastic, scalar sequences. In order to accommodate the asymmetric relation between stock returns and volatility changes, the value of  $g(z_t)$  must be a function of both the magnitude and the sign of  $z_t$ . This is done by making  $g(z_t)$  a linear combination of  $z_t$  and  $|z_t|$ :

$$g(z_t) \equiv \theta z_t + \gamma [|z_t| - E|z_t|], \quad (4.12)$$

$$z_t = \varepsilon_t / \sqrt{\sigma_t}. \quad (4.13)$$

The two components of  $g(z_t)$  are  $\theta z_t$  and  $\gamma[|z_t| - E|z_t|]$ , each with mean zero. The term  $\gamma[|z_t| - E|z_t|]$  determines the size effect, while the term  $\theta z_t$  determines the sign effect of innovations. If  $z_t$ 's distribution is symmetric, the two components are statistically independent. Over the range  $0 < z_t < \infty$ ,  $g(z_t)$  is linear in  $z_t$  with slope  $\theta + \gamma$ , and over the range  $-\infty < z_t \leq 0$ ,  $g(z_t)$  is linear with the slope  $\theta - \gamma$ . Thus,  $g(z_t)$  allows the variance process  $\{\sigma_t^2\}$  to respond asymmetrically to rises and falls in commodity prices.

If it is then assumed that  $\gamma > 0$  and  $\theta = 0$ , the innovation in  $\ln(\sigma_{t+1}^2)$  would be positive when the magnitude of  $z_t$  is larger than its expected value. Similarly, the innovation in  $\ln(\sigma_{t+1}^2)$  would be negative when the magnitude of  $z_t$  is smaller than its expected value. If, on the other hand  $\gamma = 0$  and  $\theta < 0$ , the innovation in conditional variance would be positive when returns innovations are negative, and vice versa. The exponential form of GARCH thus remedies the inability of the GARCH(1,1) model to explain the positivity and negativity of unexpected excess returns.

Nelson (1991:350–351) points out that equations 4.11 and 4.12 can solve the second limitation. This is because there are no inequality constraints in these equations and thus the  $\beta_k$  terms can be both negative and positive.

Nelson solves the third drawback by amending the variance process to take the form of  $\ln(\sigma_t^2)$  in the exponential GARCH model. Since  $\ln(\sigma_t^2)$  is a linear process, its stationarity (covariance or strict) and ergodicity are easily checked. If the shocks to  $\{\ln(\sigma_t^2)\}$  die out quickly enough, and if the component  $\{\alpha_t\}$  is removed,  $\{\ln(\sigma_t^2)\}$  is strictly stationary and ergodic. These stationarity and ergodicity criteria are exactly the same as for a general linear process with finite innovations variance, thereby solving the third drawback.

A further discussion of GARCH will be given in chapter 5, when the regression models are constructed for the testing of seasonal patterns. Section 4.5 now moves away from the observable (deterministic) approach to discuss STA, an unobserved (stochastic) approach to time series analysis.

## 4.5 An Introduction to STA

By making use of STA it is possible to test for hidden seasonal effects in data. These models are typically known as unobserved component modelling in the literature. A typical structural time series model can take the form of (Harvey & Scott, 1994:1327):

$$y_t = \mu_t + \varphi_t + \gamma_t + \varepsilon_t, \quad (4.14)$$

where  $\mu_t$  is the trend,  $\varphi_t$  is the cyclical component,  $\gamma_t$  the seasonal and  $\varepsilon_t$  the random component. Both  $\mu_t$  and  $\gamma_t$  are stochastic, but might be deterministic in limiting cases. The random component  $\varepsilon_t$  is assumed to be white noise and therefore stationary. The cyclical component  $\varphi_t$  is assumed to be stationary, while the trend component  $\mu_t$  is normally stationary in the first and second differences. The seasonal component  $\gamma_t$  is stationary when multiplied by the seasonal summation operator:

$$S(L) = I + L + L^2 + \dots + L^{s-1}, \quad (4.15)$$

where  $S(L)$  contains both real and complex unit roots,  $L$  is the lag operator, and  $s$  is the number of seasons (Harvey & Scott, 1994:1327).

The trend component  $\mu_t$  in equation 4.14 can be modelled in two ways. In the first model, the Random Walk (RW) model, the trend remains roughly constant throughout the life of the series without any continual upward or downward drift. The RW model is given by (SAS, 2006a:1631):

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2). \quad (4.16)$$

Note that if  $\sigma_\eta^2 = 0$  the model becomes  $\mu_t$ , and therefore constant. In the second model the trend is modelled as a locally linear time trend (LLT). In this model the trend is locally linear, consisting of both the level and slope. The model for  $\mu_t$  is (SAS, 2006a:1631):

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2), \quad (4.17)$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim \text{i.i.d. } N(0, \sigma_\zeta^2), \quad (4.18)$$

where the disturbance terms  $\eta_t$  and  $\zeta_t$  are assumed to be independent. The trend component  $\mu_t$  is a random walk (the level) with a drift factor  $\beta_t$ , which is the slope of the trend and follows a first order autoregressive process as represented by equation 4.17. This variation in slope and level is governed by the variances of the disturbance terms  $\eta_t$  and  $\zeta_t$  in their respective equations. The above representation of the trend is a general one that can be expressed as a random walk with drift. If both the variances are set to zero the resulting model is the deterministic linear time trend (Al-Deehani, 2006:473):  $\mu_t = \mu_0 + \beta_0 t$ . If, however,  $\sigma_\eta^2 = 0$  while  $\sigma_\zeta^2 \neq 0$ , the process will have a trend that changes relatively smoothly.

The cyclical component can be expressed as either of two cycles: deterministic or stochastic. A deterministic cycle  $\psi_t$  with frequency  $\lambda$ ,  $0 < \lambda < p$ , can be written as (Koopman, Harvey, Doornik & Shephard, 2006:62):

$$\psi_t = \alpha \cos(\lambda t) + \beta \sin(\lambda t). \quad (4.19)$$

In the case where the argument  $t$  is measured on a continuous scale,  $\psi_t$  is a periodic function with period  $2\pi / \lambda$ , amplitude  $\sqrt{(\alpha^2 + \beta^2)}$ , and phase  $\tan^{-1}(\beta / \alpha)$ . However, if  $\psi_t$  is measured only at the integer values, it will not be exactly periodic, unless  $\lambda = (2\pi j) / k$  for integers  $j$  and  $k$ .

The periodic pattern of any complexity can be written as a sum of pure cycles of different amplitudes and frequencies. This can be achieved by superimposing cycles of

different periods and amplitudes. In time series situations it is helpful to generalise the simple cyclical pattern to a stochastic cycle. This cycle will have a fixed period but time-varying amplitude and phase. Such a stochastic cycle can be given by the following recursive formula (SAS, 2006a:1632):

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix}, \quad (4.20)$$

starting with  $\psi_0 = \alpha$  and  $\psi_0^* = \beta$ . Note that  $\psi_t$  and  $\psi_t^*$  satisfy the relation  $\psi_t^2 + \psi_t^{*2} = \alpha^2 + \beta^2$  for all values of  $t$ .

In order to attain a stochastic generalisation of the cycle  $\psi_t$ , a random noise can be added to this recursion. It is also possible to add additional modelling flexibility to the model by introducing a damping factor  $\rho$ . Such a model can be given by Al-Saad and Moosa (2005:66):

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} v_t \\ v_t^* \end{bmatrix}, \quad (4.21)$$

where  $0 \leq \rho \leq 1$ , and the disturbances  $v_t$  and  $v_t^*$  are independent  $N(0, \sigma_v^2)$  variables. The stationarity properties of the random sequence  $\psi_t$  depend on the damping factor  $\rho$  (Koopman *et al.*, 2006:62). If  $\rho < 1$ ,  $\psi_t$  has a stationary distribution with zero mean and variance  $\sigma_v^2 / (1 - \rho^2)$ . If  $\rho = 1$ ,  $\psi_t$  is non-stationary.

As was the case with the cyclical component, the seasonal component can also be written as a sum of pure cycles of different amplitudes and frequencies by superimposing cycles of different periods and amplitudes. Seasonal fluctuations are a common source of variation in time series data. These fluctuations occur because of the regular changes in seasons or some other periodic events. These seasonal effects can be thought of as ‘corrections’ to the general trend of the series attributable to the seasonal variations. Because these effects sum to zero when summed over the full season cycle, the seasonal component  $\gamma_t$  is modelled as a stochastic periodic pattern of an integer

period  $s$  such that the sum  $\sum_{i=0}^{s-1} \gamma_{t-i}$  is always zero in the mean. The seasonal component can be modelled in either a dummy variable form or trigonometric form. In dummy variable form, the seasonal component can be given as (Harvey & Scott, 1994:1334):

$$\sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \quad \omega_t \sim \text{i.i.d. } N(0, \sigma_\omega^2). \quad (4.22)$$

In the case where  $\gamma_t$  is modelled in its trigonometric form as a sum of cycles of different frequencies, the model is given as (Harvey & Scott, 1994:1328):

$$\sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t}, \quad (4.23)$$

where  $\lfloor s/2 \rfloor$  equals  $s/2$  if  $s$  is even and  $(s-1)/2$  if  $s$  is odd. The cycles  $\gamma_{j,t}$  have frequencies  $\lambda_j = 2\pi j/s$ , and are specified by the matrix equation (SAS, 2006a:1633):

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \psi_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad (4.24)$$

where the disturbances  $\omega_{j,t}$  and  $\omega_{j,t}^*$  are assumed to be independent for fixed  $j$ ,  $\omega_{j,t}$  and  $\omega_{j,t}^* \sim N(0, \sigma_\omega^2)$ . In cases where  $s$  is even, the equation for  $\gamma_{s/2,t}^*$  is superfluous, and  $\gamma_{s/2,t}$  is given by SAS (2006a:1633):

$$\gamma_{s/2,t} = -\gamma_{s/2,t-1} + \omega_{s/2,t}. \quad (4.25)$$

The cycles  $\gamma_{j,t}$  are known as ‘harmonics’. In the case where the seasonal variable is deterministic, the decomposition of the seasonal effects into these harmonics is identical to its Fourier decomposition. In such a case, the sum of squares of the seasonal factors will be equal to the sum of squares of the amplitudes of the harmonics. If the disturbance variance  $\sigma_\omega^2$  is equal to zero, both the dummy variable and the trigonometric forms of seasonal components are reduced to constant seasonal effects, meaning that the seasonal component reduces to a deterministic function that is

determined by its first  $s-1$  values (SAS, 2006a:1634). Both the trigonometric- and the dummy variable-type seasonal components defined above can be considered saturated seasonal components, putting no restrictions on the  $s-1$  seasonal values.

Although STA is ideal for testing stochastic patterns, it is possible to manipulate the underlying components in order to test for deterministic effects. These tests will be discussed in chapter 5, where STA will be discussed with reference to testing for unobserved seasonality in agricultural commodity time series data.

## 4.6 Summary

There are two basic approaches when testing for seasonality in data, the deterministic approach and the stochastic approach. The models discussed under the deterministic approach include the ARMA and ARIMA models, and the GARCH models. UCM, on the other hand, allows for a stochastic approach.

Although the applications of models such as ARMA and ARIMA are numerous in the literature, STA allows tests for hidden seasonal patterns in the data, and adds value as an additional tool for the trader. While all of the models introduced in this chapter will be used to test for seasonal patterns in the data, they will also be applied in assembling a pairs portfolio in chapter 6.

Chapter 5 will explore seasonal effects in agricultural time series data by making use of the methodology discussed in this chapter. These seasonal patterns will be utilised for selecting pairs in agricultural commodities data in chapter 6, together with more traditional pairs selection tools. Chapter 6 will expand on the utilisation of the models introduced in this chapter in order to test for seasonality.

# CHAPTER 5

## Testing for Seasonality

*The divergence between expectations and outcomes provides the key to understanding history – and I interpret financial markets as a historical process*

George Soros  
(Soros, 2003:18)

### 5.1 Introduction

For the South African commodities pairs trader to be successful, he needs to be well informed about all the commodities to be include in his pairs portfolio. This knowledge should typically cover on both the fundamental and technical analysis.<sup>26</sup> Fundamental analysis includes knowledge of the commodity's price, the price of substitute and complementary products, current stock levels, surpluses from the previous season and weather patterns. Technical analysis includes information regarding the relative strength of the market, overbought and oversold conditions, the directions of trends, and trend reversals. This information will aid the pairs trader in making informed predictions regarding the current prices of the commodities; and therefore, enable the trader to guess more accurately when prices are over or under valued.

However, for the pairs trader, this information alone is not enough; it is also important to know whether any of these commodities follow any seasonal patterns. For this reason, the chapter determines whether the data used for the assembling of a pairs portfolio contain any seasonal patterns. Two forms of tests are employed to this effect: deterministic and stochastic models. Several GARCH and ARIMA models are used to test for *monthly* effects (deterministically), while unobserved component modelling are employed to test for any hidden seasonal patterns (stochastically). These tests are used as a gauge to explain price movements at the hand of fundamental analysis.

The objective of this chapter is thus to establish whether there are seasonal patterns in the agricultural commodities data. Information regarding the seasonal patterns in the

---

<sup>26</sup> Please refer to chapter 2 regarding fundamental analysis, and chapter 3 for a discussion on technical analysis.

agricultural commodities data was used in two ways. It is used, which is its traditional use, as a determinant of what the basic price trend should be at a particular point in time, and in the second way as an additional pairs selection tool.

Making use of seasonality tests in this manner is a new approach to identifying and trading pairs. As such, pairs are selected on the basis that their underlying seasonal patterns share a strong enough statistical relationship for the trader to assume that this relationship will continue in future. The more traditional methods of pairs identification includes making use of methods such as Pearson correlation, Spearman rank correlation and Beta. These methods are discussed in more detail in chapter 6.

Please note that none of these methods are used to estimate future price movements. These tests only confirm whether or not there is a strong enough statistical relationship between two series to assume that this relationship will continue, and that if prices move in opposite directions, they will revert back to a common mean.

Chapter 5 commences with a discussion of the data used, followed by a discussion on and tests for a seasonal monthly effect in all of the data, in section 5.2. This is achieved by making use of several GARCH models, which is detailed in section 5.3 below. Thereafter, in section 5.4 the implementation and results of ARIMA-based models are discussed. Section 5.5 concludes this chapter with a discussion of UCM as test for any hidden seasonal patterns in the data.<sup>27</sup>

## 5.2 Data

### 5.2.1 Introduction

The data used for the testing of pairs are the agricultural commodities most frequently traded on the SAFEX. This section gives an exposition of the data that were used in the pairs selection process. Because several data series are used, they are discussed separately.

---

<sup>27</sup> It is to be noted that tests were only initially performed on the white and yellow maize spot series. Tests were only extended to the rest of the maize series if the results on these initial series indicated that seasonality is present.

### 5.2.2 The Data

The data consists of the daily (end-of-the-day) logarithmic prices for a period stretching from the 26th of March 1998 until the 31st of October 2006 for both white and yellow maize (SAFEX, 2007). There are 2142 observations for both white and yellow maize. Although sunflower, soybean and wheat contracts are also well represented on SAFEX, tests were not performed for these contracts, because they sometimes suffer from very low liquidity levels.<sup>28</sup> This might be problematic if the pairs trader needs to execute his position while the one leg of his pairs portfolio is illiquid.

The maize data were adapted to derive the following price series: white maize spot prices (WMSPOT), white maize 1 futures contract away (WM 1 away), white maize 2 futures contracts away (WM 2 away), white maize 3 futures contracts away (WM 3 away), yellow maize spot prices (YMSPOT), yellow maize 1 futures contract away (YM 1 away), yellow maize 2 futures contracts away (YM 2 away), and yellow maize 3 futures contracts away (YM 3 away).

Even though the price difference between trading days can often be very large when switching from one month contract to the next, these series were not adjusted to smooth out these price differences. The reason for not adjusting the maize series, is that this allows for a more true reflection of the maize prices that the trader will work with.

All of the adapted maize series were transformed. To test for monthly effects in the data all the data series used for testing with ARIMA and UCM were transformed from daily values to monthly values, by taking the natural logarithm of the monthly average of every month's daily closing prices. These monthly prices are thus given by:

$$P_t = \ln \frac{1}{n} \sum_{i=1}^n Price_t, \quad (5.1)$$

where  $Price_t$  is the daily closing price of the commodity on day  $t$ . In testing for seasonality, by means of GARCH, logarithmic returns rather than prices were used. Returns ( $R_t$ ) for each agricultural commodity on day  $t$ , are expressed in the local

---

<sup>28</sup> Please refer to Tables 2.3 – 2.5 in sections 2.2.3 – 2.2.5 for the volumes traded on these commodities.

currency (rand). For the purpose of this thesis returns were calculated as the first differences in the natural logarithms of the SAFEX closing prices for the specific commodity multiplied by 100:

$$R_t = \ln\left(\frac{Price_t}{Price_{t-1}}\right) \times 100, \quad (5.2)$$

where  $R_t$  is the return for the period  $t$ ,  $Price_t$  is the daily closing prices at time  $t$ , and  $t$  is the time measured in days. Each return series is expressed as a percentage.<sup>29</sup>

### 5.2.3 Conclusion

Monthly logarithmic prices were used for the ARIMA and UCM models. However, in testing for monthly seasonality with GARCH daily logarithmic returns were used. The next section covers the development of the GARCH model, which was used to investigate the presence of monthly seasonal effects.

## 5.3 Modelling for Seasonal Patterns with GARCH

### 5.3.1 Introduction

GARCH is frequently used in the literature to test for deterministic seasonality (these studies include French *et al.*, 1987; Akgiray, 1989; and Nelson, 1991). In this section, the GARCH model that was developed to test for monthly effects in the maize data used in the pairs construction process is presented.<sup>30</sup>

The final model for each data series was derived by making use of the literature that was discussed above, as well as various SAS<sup>®</sup> procedures (SAS, 2007a). It is to be noted that other models were also tested to ensure objectivity. In order to ascertain which one best models monthly effects the models were tested by means of various statistical measures to ascertain their statistical appropriateness. In testing for monthly seasonality the monthly spot series for white maize were used first, followed by tests using yellow maize data.

<sup>29</sup> Modelling data in terms of returns is typical in the literature, see for example Nelson (1991) and also Mills, Siringopoulos, Markellos, and Harizanis (2000).

<sup>30</sup> The term 'maize' refers to both white and yellow maize.

### 5.3.2 Methodology

In testing for monthly seasonal effects in agricultural commodities time series data a variety of models can be employed (the selection process regarding the model is discussed in section 5.3.2.1.2 below). The same type of model should be used for all the series tested.<sup>31</sup> A possible approach is to allow for the inclusion of all twelve months in the model as dummy variables. In such a case a restriction on the parameter estimates has to be included to avoid the dummy variable trap.<sup>32</sup> This model can be given as (Keim & Stambaugh, 1984):

$$R_t = \alpha_0 + \alpha_1 Jan_t + \alpha_2 Feb_t + \alpha_3 Mar_t + \dots + \alpha_{12} Dec_t + \varepsilon_t, \quad (5.3)$$

with restriction  $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 + \alpha_8 + \alpha_9 + \alpha_{10} + \alpha_{11} + \alpha_{12} = 0$ .

Market return is given by  $R_t$  while  $Jan_t$ ,  $Feb_t$ ,  $Mar_t$ ,  $Apr_t$ ,  $May_t$ ,  $Jun_t$ ,  $Jul_t$ ,  $Aug_t$ ,  $Sept_t$ ,  $Oct_t$ ,  $Nov_t$ , and  $Dec_t$  are dummy variables for January through December at time  $t$ , and  $\varepsilon_t$  represents the error term. For the monthly dummy variables  $Jan_t = 1$  if month  $t$  is January and 0 otherwise,  $Feb_t = 1$  if the month  $t$  is February and 0 otherwise,  $Mar_t = 1$  if month  $t$  is March and 0 otherwise; and so on.

Another approach would be to allow for one of the months to be used as a benchmark month. In such a case the dummy variable also takes a value of unity for a given month and a value of zero for all other months, but the coefficient of each dummy variable measures the incremental effect of that month relative to the benchmark month. This model can be given by Gibbons and Hess (1981):

$$R_t = \alpha_0 + \sum_{i=1}^{11} \alpha_i Month_{it} + \varepsilon_t, \quad (5.4)$$

where any month can be used as the benchmark month. Equation 5.4 above also includes an intercept term ( $\alpha_0$ ) along with dummy variables for all months except one.

In the above model,  $Month_{it}$  is the dummy variable that is equal to 1 for month  $i$  and 0

<sup>31</sup> When testing for monthly seasonality in the white maize, and yellow maize data, the same type of model was used in terms of the amount of dummy variables included in the model.

<sup>32</sup> The dummy variable trap is a situation of perfect colinearity or perfect multi-colinearity, should there be more than one exact relationship between the variables (Gujarati, 2003: 302).

otherwise. Following Gibbons and Hess (1981), and Berument and Kiyamaz (2003:367) this latter approach was used to test for a monthly effect in the South African maize data. To test for a month effect in the South African maize data, each of the months was alternately used as benchmark month.

The model for each data series (white maize and yellow maize) was tested for both autocorrelation and heteroskedasticity. Testing for autocorrelation was done by making use of the *DWPROB*-option in the PROC AUTOREG procedure of SAS<sup>®</sup> (SAS, 2006b).

In the case of equation 5.4 above, the Durbin-Watson statistics for higher-order autocorrelation were used to test for autocorrelation in the monthly ordinary least square (OLS) residuals in all the data series. It is advisable to specify an order at least as large as the order of any potential seasonality, since seasonality produces autocorrelation at the seasonal lag (SAS, 2006b).

For equation 5.4,  $DW = 12$  was specified for monthly data because the largest expected autocorrelation is twelve months for a monthly effect. However, this test does not specify which autoregressive orders should be tested, but only indicates that autocorrelation of this (or higher) orders are present in the log returns data. Heteroskedasticity was simultaneously tested by making use of the *ARCHTEST*-option in the PROC AUTOREG procedure of SAS<sup>®</sup> (SAS, 2006b). More regarding the tests for autocorrelation and heteroskedasticity follows later in section 5.3.2.1.2.

Identical tests were performed for the white maize and yellow maize data. Sections 5.3.2.1 and 5.3.2.2 below discuss the various models regarding each of the maize series and the results of the tests for autocorrelation and heteroskedasticity. Certain assessment of fit tests regarding the goodness-of-fit of all the models was also performed in order to ascertain whether the model used, was indeed the best model. These results are also discussed during the model selection process below.

It is to be noted that tests for seasonality were first performed on the WMSPOT and YMSPOT series. If these results indicated that GARCH was successful in identifying

the seasonal patterns in these two series, all of the other series are also tested by means of the appropriate GARCH models.

### 5.3.2.1 White Maize

#### 5.3.2.1.1 Data Results

Prior to discussing the results of the model employed in testing for monthly seasonality in the white maize spot series, the data are briefly described. Following the description of the data, the results regarding the tests for autocorrelation and heteroskedasticity are discussed in section 5.3.2.1.2.

Table 5.1 below reports on the preliminary descriptive statistics for both the returns of the entire study period (26th of March 1998 until the 31st of October 2006) and the returns for the month. The first column in Table 5.1 reports the monthly mean returns, median, maximum and minimum values, standard deviation, skewness and kurtosis measures for all months. The mean returns for the entire study period is 0.005 with a standard deviation of 0.101. Table 5.1 also reports the descriptive statistics for the return series of each month of the year.

The findings regarding the returns of every month indicate that the mean returns for November are the highest (0.065) followed by December, with mean returns of 0.047, while February has the lowest mean returns of -0.056. Negative skewness is present in the data for January, April, May, July, August and December; and positive skewness for February, March, June, September, October, and November.

TABLE 5.1 SUMMARY STATISTICS: WHITE MAIZE LOG RETURNS PRICES

	All Months	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
<b>No of obs.</b>	102	8	8	8	8	9	9	9	9	9	9	8	8
<b>Mean</b>	0.005	0.018	-0.056	0.001	-0.035	-0.019	-0.003	-0.002	0.001	0.021	0.018	0.065	0.047
<b>Median</b>	0.005	0.003	0.006	-0.016	-0.036	-0.025	0.001	-0.029	0.021	0.014	-0.004	0.080	0.016
<b>Maximum</b>	0.310	0.178	0.094	0.310	0.031	0.100	0.115	0.141	0.081	0.440	0.134	0.194	0.193
<b>Minimum</b>	-0.335	-0.175	-0.335	-0.210	-0.118	-0.133	-0.114	-0.119	-0.085	-0.076	-0.040	-0.061	-0.197
<b>Std. Dev</b>	0.101	0.130	0.151	0.173	0.050	0.074	0.066	0.100	0.065	0.077	0.057	0.036	0.136
<b>Skewness</b>	-0.151	-0.186	-1.112	0.507	-0.300	0.136	0.120	-0.288	-0.119	1.298	1.221	0.021	-0.489
<b>Kurtosis</b>	1.229	-1.203	0.125	0.159	-0.348	-0.171	0.643	-1.327	-2.026	2.510	0.867	1.353	-0.133

The data suggest that there might be a positive November effect in mean returns and a negative February effect in mean returns. In order to properly assess the accuracy of this supposition, the data is simulated with a GARCH model. The model selection process is discussed in the next section.

### 5.3.2.1.2 Model Selection

In order to select the final model that was used in testing for monthly seasonality in the white maize data, tests for the existence of autocorrelation in the data was first performed. According to the first-order Durbin-Watson test the hypothesis of first-order autocorrelation in the white maize spot series was statistically significant at the 5 percent level of significance, with  $p < .0001$ .<sup>33</sup> Since the first-order Durbin-Watson test was found to be statistically significant, the tests for the higher orders could be ignored. The above results call for autocorrelation correction. In order to address the issue of autocorrelation Berument and Kiyamaz (2001:3) is followed by inserting lags in the return equation.

Autocorrelation was found to be particularly severe in emerging markets data because of their low level of liquidity (De Santis & Imrohoroglu, 1997). This explains the existence of autocorrelation in the white maize data. In order to decide on the number of autoregressive orders that were used, the stepwise autoregressive process was performed by making use of the Yule-Walker method (Maddala, 2002:521-522). When using stepwise autoregression, an *NLAG*-option value larger than the order of any potential seasonality should be used, since seasonality produces autocorrelation at the seasonal lag (SAS, 2006b). An initial order of 13 was used. This process was performed by making use of the *BACKSTEP*-option in SAS<sup>®</sup>'s PROC AUTOREG procedure (SAS, 2006b). The test revealed that only the first lag was statistically significant at the 1 percent level of significance.<sup>34</sup> The results of the test thus indicate that an AR (1) process could be used to test for seasonality in the white maize data.

In testing for heteroskedasticity the *ARCHTEST*-option in the PROC AUTOREG procedure of SAS<sup>®</sup> was used (SAS, 2006b). One of the key assumptions of an ordinary

---

<sup>33</sup> See statistical report 1.1 in the appendix for the full report on the autocorrelation test on the OLS model before lags.

<sup>34</sup> See statistical report 1.2 in the appendix for the full report on the estimation of the autoregressive lags.

regression model is that the errors have the same variance throughout the sample, and is thus homoskedastic. In the case where the error variance is not constant, the data is said to be heteroskedastic (SAS, 2006b).

The Q statistics that test for changes in variance across time were performed using lag windows ranging from 1 through 12. These tests strongly indicate heteroskedasticity, with  $p < 0.1$  for the first eight lag windows.<sup>35</sup> The Lagrange multiplier (LM) tests also indicate heteroskedasticity, and  $p$  values of less than 0.1 were found for the first three lag windows. In this case, heteroskedastic errors were eliminated by modelling the variance of the returns with Nelson's (1991) EGARCH(1,1) model.

The final model thus took the form of an AR(1) EGARCH(1,1) model, this is discussed in section 5.3.2.1.3. This is followed by a discussion of the model results for the white maize data. This model was derived from various studies in the literature as well as the empirical research undertaken and detailed above.

#### 5.3.2.1.3 The Model

This project made use of a GARCH model to account for both the autocorrelation and heteroskedasticity found in the white maize data. Nelson (1991) and Ogum *et al.* (2002) are followed in allowing for the model to respond asymmetrically to the increases and decreases in the price; and Engle *et al.* (1987) were followed in making use of 'in-mean' (risk factor) effects. The risk factor included in the model allows for modelling the additional risk (if any) traders take when seeking to earn higher returns. As previously mentioned, Nelson (1991) introduced EGARCH to counter the asymmetric conditional variance behaviour of returns brought about by the 'leverage effect' (Black, 1976).<sup>36</sup>

Nelson's (1991) EGARCH( $p,q$ ) model captures skewness and asymmetry.<sup>37</sup> The final model takes the form of an AR(1) EGARCH(1,1)-M model and is given by:

---

<sup>35</sup> See statistical report 1.3 in the appendix for a report on the heteroskedasticity in the white maize data.

<sup>36</sup> According to Black's (1976) 'leverage effect', financial leverage increases as the market value of equity falls, causing conditional volatility to increase. Financial leverage similarly decreases as the market value of equity rises, causing conditional volatility to decrease.

<sup>37</sup> See section 4.4 for a discussion on GARCH, and particularly EGARCH.

$$R_t = \alpha_0 + \phi R_{t-1} + \alpha_2 Jan_t + \alpha_3 Feb_t + \alpha_4 Mar_t + \dots + \alpha_{13} Dec_t + \delta \sigma_t + \varepsilon_t, \quad (5.5)$$

$$\varepsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2), \quad (5.6)$$

$$\ln(\sigma_t^2) = \varphi_0 + \varphi_1 \left\{ |\varepsilon_{t-1} / \sigma_{t-1}| - \sqrt{2/\pi} \right\} + \gamma \varepsilon_{t-1} / \sigma_{t-1} + \beta \ln(\sigma_{t-1}^2), \quad (5.7)$$

where equation 5.5 is the return equation,<sup>38</sup> equation 5.6 indicates how the error terms are distributed and equation 5.7 is the variance equation. In equation 5.5,  $R_t$  represents the measure of market return,  $Jan_t$ ,  $Feb_t$ ,  $Mar_t$ ,  $Apr_t$ ,  $May_t$ ,  $Jun_t$ ,  $Jul_t$ ,  $Aug_t$ ,  $Sept_t$ ,  $Oct_t$ ,  $Nov_t$ , and  $Dec_t$  are dummy variables for January through December at time  $t$ .  $Jan_t = 1$ , if month  $t$  is January and 0 otherwise,  $Feb_t = 1$ , if the month  $t$  is February and 0 otherwise,  $Mar_t = 1$ , if month  $t$  is March and 0 otherwise, and so forth.  $(R_t)$  is considered to be linearly related to the previous months' returns as well as its own standard deviation.

The component  $(\phi R_{t-1})$  allows for the returns in the present month to be influenced by the returns of one period (month) prior. The stochastic disturbance (error) term is indicated by the variable  $(\varepsilon_t)$ . The error term  $(\varepsilon_t)$  is assumed to follow a normal distribution with  $\Omega_{t-1}$  being the set of relevant information available at time  $t$ .

Following Engle *et al.* (1987), the equation also allows for a risk factor in the form of the 'in-Mean' parameter  $(\delta)$ . This parameter is introduced in order to determine whether or not investors are rewarded for their exposure to market risk. According to the CAPM mean-variance hypothesis, large standard deviations (variances or volatility) are expected to be associated with large returns (Litterman, 2003:37). It follows that  $(\delta)$  is expected to be greater than zero. Consequently, the parameter  $(\delta)$  determines the relationship between returns and volatility.

---

<sup>38</sup> Note that this model is implemented twelve times, making use of a different reference month every time in the returns equation. This is done to avoid the dummy variable trap discussed above.

The conditional variance equation 5.7 follows an EGARCH(1,1) process (Nelson, 1991), which allows for time-varying heteroskedasticity in the errors. The AR(1) EGARCH(1,1)-M model is more general than the standard GARCH model, in that it allows innovations of different signs to have a differential impact on volatility and allows bigger shocks to have a larger impact on volatility. This model is further improved by modelling the logarithm of the conditional variance ( $\sigma_t$ ), eliminating the need to restrict parameter values to avoid negative variances.

In equation 5.7 the parameter ( $\phi_1$ ) measures the impact of innovations on conditional volatility at time  $t$ . The parameter ( $\gamma$ ) permits the asymmetric response of conditional variance to innovations of a differing sign (positive or negative). In the case of ( $\gamma$ ) in equation 5.7 being negative, negative realisations of the innovation (risk factor) in equation 5.5 will generate more volatility than positive realisations will. If, however, ( $\gamma$ ) is positive, negative realisations of the innovation in equation 5.5 will generate less volatility than positive realisations will.

The presence of the leverage effect can be tested by the assumption that  $\gamma < 0$ . The impact is asymmetric in the case where  $\gamma \neq 0$  and the most recent residual term impact is exponential, rather than quadratic. ‘Good’ news ( $\varepsilon_{t-1} > 0$ ) will have an impact of  $(\phi + \gamma)/\sigma_{t-1}$  while ‘bad’ news ( $\varepsilon_{t-1} < 0$ ) will have an impact of  $(\phi - \gamma)/\sigma_{t-1}$  (Ogum *et al.*, 2002:11). The parameter ( $\beta$ ) is the autoregressive term on lagged conditional volatility, thus reflecting the weight given to the previous period’s conditional volatility in the conditional volatility at time  $t$ . It measures the persistence of shocks to the conditional variance.

#### 5.3.2.1.4 Model Results

The AR(1) EGARCH(1,1)-M model used in this project was derived from the literature and empirically verified by making use of several SAS<sup>®</sup> procedures (SAS, 2007b). Equation 5.5 allows for lagged values of returns to influence present returns. This was done in order to remove autocorrelation from the data. In order to assure that autocorrelation was indeed eradicated from the data, the Durbin-Watson test was

performed again after using the AR(1) EGARCH(1,1)-M model. The results of the Durbin-Watson test indicate that autocorrelation was eliminated successfully for all twelve of the tests (for all the reference months).<sup>39</sup>

Another method of determining whether autocorrelation has been removed successfully is to test the standardised residuals for normal distribution. If the autocorrelation has been removed from the data, the standardised residuals will follow a normal distribution (Mukherjee, White & Wuyts, 1998:379).<sup>40</sup> To test for normality, the Jarque-Bera skewness-kurtosis test for normal distribution was used. The Jarque-Bera test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution. This statistic is given by (Quantitative Micro Software (QMS), 2004a:300):

$$JB = \frac{N-k}{6} \left( S^2 + \frac{(K-3)^2}{4} \right), \quad (5.8)$$

where  $S$  is the skewness,  $K$  is the kurtosis, and  $k$  is the number of estimated coefficients used to create the series. The  $p$ -value reported for the Jarque-Bera statistic, is the probability that a Jarque-Bera test statistic exceeds (in absolute value) the observed value under the null hypothesis. A small probability value leads to the rejection of the null hypothesis of normal distribution (QMS, 2004a:300). Table 5.2 below reports on the results of the normality tests. The first row in Table 5.2 reports the Jarque-Bera test statistic for each of the tests performed with the different reference months. The skewness and kurtosis of the standardised residual for each reference month were also included.

TABLE 5.2 NORMALITY TESTS RESULTS: STANDARDISED RESIDUALS

Normality Test	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jarque-Bera ( $p$ -value)	0.99	0.94	0.79	0.86	0.87	0.68	0.93	0.76	0.81	0.98	0.83	0.89
Skewness	0.01	-0.08	0.11	-0.07	-0.04	-0.17	-0.09	-0.18	-0.16	0.05	-0.21	0.11
Kurtosis	2.98	2.94	2.75	2.77	2.76	2.73	3.04	2.94	2.93	2.96	2.86	3.00

<sup>39</sup> See statistical report 1.4 in the appendix for the Durbin-Watson statistics.

<sup>40</sup> The standardised residual is the ratio of the residual to its standard error:  $Std. res = \frac{\varepsilon_t}{\sqrt{\sigma^2}}$

By making use of an AR(1) EGARCH(1,1)-M model, it was also possible to eliminate heteroskedasticity from the data. These results were confirmed by the Q statistics and LM multiplier tests.<sup>41</sup> The results regarding the ‘in-Mean’ risk factor ( $\delta$ ) are all significant at the 1 percent level of significance with the exception of February. This result indicates that the incentive for risk takers in the South African market is fairly strong.

In the variance equation the parameter ( $\delta$ ) is found to be statistically significant for the tests performed with January, May, June, July, August, September, and November as reference month. However, the parameter ( $\beta$ ) is not statistically significant for the tests performed with February, March, April, October and December as reference months. The parameter ( $\phi_1$ ) is significant at the 1 percent level of significance, for all of the reference months. The parameter ( $\gamma$ ) is also statistically significant at the 1 percent level of statistical significance for all the reference months with the exception of February.

#### 5.3.2.1.5 Final Results – White Maize

Table 5.3 below reports on the various monthly effects in terms of the benchmark month in the left-hand side column. The 2 columns on the far right hand side indicate which month has the highest and lowest logarithmic returns. Only statistically significant results were included. The estimates in bold are significant at the 5 percent level of statistical significance, while the italicised estimates are significant at the 10 percent level of statistical significance.

---

<sup>41</sup> View statistical report 1.5 in the appendix for the full report on the heteroskedasticity in the data after making use of the AR(1) EGARCH(1,1)-M model was used to test for seasonality.

TABLE 5.3 MONTHLY SEASONAL EFFECTS OF WHITE MAIZE LOG RETURNS PRICES

Benchmark Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	High	Low
Jan			-0.047	-0.092	-0.065	-0.052				-0.051			Nov	Apr
Feb				-0.059						-0.027			Nov	Apr
Mar				-0.059		-0.041		-0.042		-0.041			Feb	Apr
Apr	0.118	0.026	0.068				0.041	0.039	0.050	0.043	0.005	0.096	Jan	May
May											0.062	0.041	Nov	Jun
Jun		0.042		-0.047									Feb	Apr
Jul	0.071	0.049						0.043			0.075	0.072	Nov	Apr
Aug				-0.063		-0.039	-0.028						Jan	Apr
Sep				-0.075		-0.066	-0.056			-0.029			Mar	Apr
Oct				-0.049				0.036	0.03		0.053	0.036	Nov	Apr
Nov				-0.080	-0.017	-0.034			0.025			0.025	Sep	Apr
Dec		-0.029	-0.059	-0.069	-0.063	-0.043		-0.033		-0.043			Nov	Apr

According to the GARCH tests, there is no strong positive seasonal effect present in the white maize returns series data. Six of the tests indicate a positive November effect, while two tests indicated a positive January effect, two tests indicate a positive February effect, one test indicate a positive March effect, and one test a positive September effect. A strong negative effect was recorded for April, indicating that prices should be low in April.

However, this result does not correlate with the actual white maize seasonal pattern. Traditionally, white maize prices peak around December/January, and reach their low in the autumn/winter months. The high price for white maize in December is affected by uncertainty in the market caused by the dry period during the planting season in the Western parts of South Africa (Van Zyl, 2006). The traditional low prices experienced in the winter months can be attributed to the certainty about the physical stock levels during this period.

Because GARCH failed to capture these seasonal patterns existing in the white maize data, this deterministic type model may not be effective as tool for the commodities pairs trader. In order to establish whether GARCH identifies similar monthly seasonality in the yellow maize data, however, the same procedure was repeated for the yellow maize data series. The tests for yellow maize are described in the following section.

### 5.3.2.2 Yellow Maize

#### 5.3.2.2.1 Data Results

First, a brief description of the data is given, followed by the results regarding the tests for autocorrelation and heteroskedasticity in section 5.3.2.2.2. Table 5.4 below reports the preliminary descriptive statistics in terms of the returns of the entire study period, as well as for the month. The first column in Table 5.4 reports the monthly mean returns, median, maximum and minimum values, standard deviation, skewness, and kurtosis measures for all months. For the yellow maize data, the mean returns for the entire study period is 0.007 with a standard deviation of 0.080.

Table 5.4 also reports the descriptive statistics for the return series of each month of the year. The results of the preliminary testing revealed that the mean returns for November are the highest (0.049) followed by September and December with mean returns of 0.044 each, while April has the lowest mean returns of -0.051 followed by March (-0.028). Negative skewness is present in the data for January, February, April, June, July, November and December; and positive skewness for March, May, August, September, and October. No excess kurtosis is present for returns in any of the individual months. The data, thus, suggest that there might be a positive November effect and a negative April effect in mean returns. The data were tested with a GARCH model to assess whether these effects do exist. The model selection process is discussed in section 5.3.2.2.2 below.

TABLE 5.4 SUMMARY STATISTICS: YELLOW MAIZE LOG RETURNS PRICES

	All Months	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
<b>No of obs.</b>	102	8	8	8	8	9	9	9	9	9	9	8	8
<b>Mean</b>	0.007	0.031	-0.027	-0.028	-0.051	-0.009	-0.003	-0.001	0.004	0.044	0.026	0.049	0.044
<b>Median</b>	0.010	0.032	-0.013	-0.056	-0.045	-0.007	0.014	0.031	-0.009	0.034	0.002	0.050	0.037
<b>Maximum</b>	0.246	0.174	0.074	0.246	0.009	0.090	0.116	0.091	0.068	0.132	0.149	0.162	0.186
<b>Minimum</b>	-0.166	-0.152	-0.166	-0.158	-0.133	-0.104	-0.109	-0.144	-0.050	-0.012	-0.047	-0.055	-0.116
<b>Std. Dev</b>	0.080	0.103	0.093	0.130	0.044	0.063	0.064	0.083	0.049	0.044	0.059	0.075	0.092
<b>Skewness</b>	0.104	-0.475	-0.334	1.483	-0.687	0.147	-0.256	-0.758	0.278	0.879	1.134	-0.003	-0.273
<b>Kurtosis</b>	0.152	0.036	-1.650	2.390	0.563	-0.567	0.995	-0.843	-1.981	0.916	1.351	-0.757	0.444

### 5.3.2.2.2 Model Selection

As was the case with the white maize data, tests for autocorrelation in the data were first performed. The Durbin-Watson test for first-order autocorrelation in the yellow maize spot series is statistically significant at the 1 percent level of statistical significance with a *p-value* of  $< .0001$  for the hypothesis of first-order positive autocorrelation in the yellow maize spot series data.<sup>42</sup> Since the first-order Durbin-Watson test is found to be significant, the tests for the higher orders were ignored. To address autocorrelation, lags were inserted in the return equation.

The number of autoregressive orders that was used was determined by performing the stepwise autoregressive process in SAS<sup>®</sup> by making use of the *BACKSTEP*-option in SAS<sup>®</sup>'s PROC AUTOREG procedure (SAS, 2006b).

The test results revealed that lags (1 and 12) are statistically significant at the 5 percent level of significance.<sup>43</sup> The results of the test thus indicate that an AR (1 or 12) process could be used to test for seasonality in the yellow maize data. In order to decide on the best model (with the correct lag length), Schwarz's Bayesian information criterion (SBC) was used (Maddala, 2002:488). The Akaike Information criterion (AIC) and the SBC are computed as follows (Hatanaka, 1996:227):

$$\text{AIC} = -2\ln(L) + 2k, \quad (5.9)$$

$$\text{SBC} = -2\ln(L) + \ln(N)k, \quad (5.10)$$

where  $L$  is the value of the likelihood function evaluated at the parameter estimates,  $N$  is the number of observations, and  $k$  is the number of estimated parameters (Chatfield, 2004:256). The model with the lowest AIC and SBC values is the one that expresses the patterns in the data the best. The results of these tests (of the different lag length models) are compared after establishing whether heteroskedasticity is present.

---

<sup>42</sup> See statistical report 1.6 in the appendix for the full report on the autocorrelation test on the OLS model before lags.

<sup>43</sup> See statistical report 1.7 in the appendix for the full report on the estimation of the autoregressive lags.

The *ARCHTEST*-option in SAS<sup>®</sup>'s PROC AUTOREG procedure was used to test for heteroskedasticity (SAS, 2006b). As was the case with the tests for monthly seasonality in the white maize data, the Q statistics that test for changes in variance across time were performed using lag windows ranging from 1 through 12. These tests did not indicate heteroskedasticity, with none of the *p*-values indicating statistical significance at the 10 percent level for any of the lag windows.<sup>44</sup> The LM tests also indicate that there is no heteroskedasticity in the data.

In order to compare the white maize results with that of the yellow maize, the yellow maize data was also be modelled by making use of Nelson's (1991) EGARCH(1,1) model. For the yellow maize data, the SBC results revealed that using 1 autoregressive lag would be the best way to model the yellow maize data. The final model thus took the form of an AR(1) EGARCH(1,1) model. Section 5.3.2.2.3 below, discusses the final model used in modelling for monthly seasonal effects in the yellow maize data, which is followed by a discussion of the model results in section 5.3.2.2.4. This model was derived from various studies in the literature as well as the empirical research undertaken and detailed above.

### 5.3.2.2.3 The Model

As was the case with the white maize data, this research project made use of an EGARCH model to account for the autocorrelation found in the yellow maize data. The final model took the form of an AR(1) EGARCH(1,1)-M model which is given by:

$$R_t = \alpha_0 + \phi R_{t-1} + \alpha_2 Jan_t + \alpha_3 Feb_t + \alpha_4 Mar_t + \dots + \alpha_{13} Dec_t + \delta \sigma_t + \varepsilon_t, \quad (5.11)$$

$$\varepsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2), \quad (5.12)$$

$$\ln(\sigma_t^2) = \varphi_0 + \varphi_1 \left\{ \left| \varepsilon_{t-1} / \sigma_{t-1} \right| - \sqrt{2/\pi} \right\} + \gamma \varepsilon_{t-1} / \sigma_{t-1} + \beta \ln(\sigma_{t-1}^2), \quad (5.13)$$

where equation 5.11 is the return equation, equation 5.12 indicates how the error terms are distributed and equation 5.13 is the variance equation. In equation 5.11,  $R_t$

---

<sup>44</sup> The complete report indicating heteroskedasticity can be viewed in statistical report 1.8, see the appendix.

represents the measure of market return,  $Jan_t$ ,  $Feb_t$ ,  $Mar_t$ ,  $Apr_t$ ,  $May_t$ ,  $Jun_t$ ,  $Jul_t$ ,  $Aug_t$ ,  $Sept_t$ ,  $Oct_t$ ,  $Nov_t$ , and  $Dec_t$  are dummy variables for January through December at time  $t$ .  $Jan_t = 1$ , if month  $t$  is January and 0 otherwise,  $Feb_t = 1$ , if the month  $t$  is February and 0 otherwise,  $Mar_t = 1$ , if month  $t$  is March and 0 otherwise, and so on.  $(R_t)$  is considered to be linearly related to the previous months' returns as well as its own standard deviation  $(\sigma_t)$ .

The component  $(\phi R_{t-1})$  allows for the returns in the present month to be influenced by the returns of one period (month) prior. The stochastic disturbance is denoted by the component  $(\varepsilon_t)$  and is assumed to follow a normal distribution with  $\Omega_{t-1}$  the set of relevant information available at time  $t$ . The equation furthermore allows for a risk factor in the form of the 'in-Mean' parameter  $(\delta)$ , while the logarithm of the conditional variance  $(\sigma_t)$  is modelled in order to eliminate the necessity to restrict parameter values in avoiding negative variances. The variance equation (5.13) for the AR(1) EGARCH(1,1)-M model is identical to equation 5.7, used in the AR(1) EGARCH(1,1)-M model above in modelling the white maize data.<sup>45</sup>

#### 5.3.2.2.4 Model Results

The AR(1) EGARCH(1,1)-M model used in testing for monthly seasonality in the yellow maize data was derived from the literature and empirically verified by making use of several SAS<sup>®</sup> procedures (SAS, 2007b). By allowing the component  $(\phi R_{t-1})$  in equation 5.11 to lag the values of returns, autocorrelation was eradicated from the data.

As was the case with the white maize tests, the Durbin-Watson test was performed in order to assure that autocorrelation had been eradicated from the data after using the AR(1) EGARCH(1,1)-M model. The results of the Durbin-Watson test indicate that autocorrelation was eliminated successfully for all twelve of the tests (for all the reference months).<sup>46</sup>

<sup>45</sup> See equation 5.7 above for the specifications of  $\varphi_1, \gamma$  and  $\beta$

<sup>46</sup> See statistical report 1.9 in the appendix for the Durbin-Watson statistics.

Normality tests were then performed on the standardised residuals. The eradication of autocorrelation from the data was tested by making use of the Jarque-Bera skewness-kurtosis test for normal distribution. Table 5.5 below reports on the results of the normality tests. The first row in Table 5.5 reports the Jarque-Bera test statistic for each of the tests performed with the different reference months. The skewness and kurtosis of the standardised residual for each reference month are also included.

TABLE 5.5 NORMALITY TESTS RESULTS: STANDARDISED RESIDUALS

Normality Test	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jarque-Bera <i>p</i> -value	0.38	0.32	0.30	0.53	0.35	0.29	0.48	0.33	0.24	0.41	0.20	0.36
Skewness	-0.20	-0.10	0.00	-0.15	-0.25	-0.22	-0.05	-0.23	-0.25	-0.25	-0.02	-0.25
Kurtosis	2.45	2.29	2.25	2.54	2.50	2.37	2.42	2.44	2.36	2.58	2.13	2.51

It was also possible to eliminate heteroskedasticity from the data by making use of an AR(1) EGARCH(1,1)-M.<sup>47</sup> The results regarding the 'in-Mean' risk factor ( $\delta$ ) are significant at the 1 percent level of significance, indicating that the incentive for risk takers in the South African market is fairly strong. In the variance equation the parameters ( $\varphi_1$ ) and ( $\gamma$ ) are significant at the 1 percent level of statistical significance.<sup>48</sup> The parameter  $\beta$  is statistically significant at the 1 percent level of statistical significance for all the reference months besides July.

### 5.3.2.2.5 Final Results – Yellow Maize

Table 5.6 below reports the various monthly effects in terms of the benchmark month in the left-hand side column. As is the case with the white maize data, only the statistically significant results are included, reporting the highest and lowest logarithmic returns on the far right-hand side. The estimates in bold are significant at the 5 percent level of statistical significance while the italicised estimates are significant at the 10 percent level of statistical significance.

<sup>47</sup> Statistical report 1.10 in the appendix reports on the elimination of the heteroskedasticity in the data after making use of the AR(1) EGARCH(1,1)-M model.

<sup>48</sup> The parameters ( $\varphi_1$ ) and ( $\gamma$ ) are statistically significant at the 5 percent level of statistical significance irrespective of the month being used as reference month.

TABLE 5.6 MONTHLY SEASONAL EFFECTS OF YELLOW MAIZE LOG RETURNS PRICES

Benchmark Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	High	Low
Jan		-0.104	-0.141	-0.110	-0.053	-0.041		-0.047		-0.032	0.029	0.012	Nov	Mar
Feb	0.111				0.051	0.061	0.066	0.098	0.135	0.087	0.084	0.105	Sep	Mar
Mar	0.145	0.076		0.054	0.098	0.131	0.152	0.127	0.168	0.113	0.180	0.146	Nov	Apr
Apr	0.116		-0.026		0.057	0.069	0.077	0.063	0.119	0.090	0.148	0.128	Nov	Mar
May	0.083	-0.024	-0.087	-0.038					0.065	0.040	0.095	0.092	Nov	Mar
Jun	0.023	-0.076	-0.113	-0.076	-0.018				0.049		0.059	0.034	Nov	Apr
Jul		-0.078	-0.099	-0.083					0.045		0.069		Nov	Mar
Aug	0.058	-0.042	-0.087	-0.062			0.016		0.063	0.018	0.078	0.071	Nov	Mar
Sep	-0.029	-0.118	-0.155	-0.122	-0.065	-0.055	-0.047	-0.061		-0.050			Nov	Mar
Oct	0.036	-0.069	-0.114	-0.086	-0.023			-0.020	0.037		0.058	0.046	Nov	Mar
Nov	-0.050	-0.150	-0.175	-0.125	-0.101	-0.092	-0.114	-0.059	-0.039	-0.052		-0.028	Dec	Mar
Dec	-0.013	-0.122	-0.176	-0.121	-0.082	-0.071	-0.056	-0.074	-0.018	-0.041			Nov	Mar

For the yellow maize data, a strong positive November effect is present. Ten of the tests indicated a positive November effect, while one test indicated a positive December effect and one test indicated a positive September effect. A strong negative effect was recorded for March (ten of the tests indicated a negative March effect), indicating that prices should be low in March. As is the case with the white maize results, this result does not correlate with the actual white maize seasonal pattern. Traditionally, similarly to white maize, yellow maize prices peak around December/January and reach their low in July/August. The high and low prices for yellow maize are caused by the same factors as those of white maize.

Because GARCH fails to capture these seasonal patterns existing in the yellow maize data, it will not be an effective tool for the commodities pairs trader. GARCH can therefore not be effective as a tool for finding pairs.

### 5.3.3 Conclusion

Should the pairs trader use an econometric model as a tool for identifying pairs, it is necessary that the model is able to explain the seasonality in the data. The GARCH-type models do not explain the seasonal effects in South Africa's agricultural commodities data adequately, and can therefore not be used as a tool for identifying commodity pairs. This model would also be ineffective in deciding when to execute trades.

Since GARCH was not successful in identifying seasonal patterns in the WMSPOT and YMSPOT series, no further tests were performed using the other maize series. In addition, GARCH was also not implemented during the pairs selection process that is detailed in chapter 6. In order to further test for seasonality in the maize data, a second deterministic model was used in the form of a X12 ARIMA-type model, which is discussed in section 5.4 below.

## 5.4 Modelling for Seasonal Patterns with X12

### 5.4.1 Introduction

The X12 (also called Census 12) procedure was developed by the Time Series Staff of the Statistical Research Division, United States Bureau of the Census, and is most commonly used to produce seasonally-adjusted series (SAS, 2006c). Eliminating the seasonal component from an economic series facilitates comparison among consecutive months or quarters; furthermore, plots of these seasonally adjusted series are often more informative about trends than plots of the unadjusted series (SAS, 2006c).

In order to produce a seasonally-adjusted series, the X11-ARIMA process was first used to separate the original series  $O_t$  into four components. The first component is the seasonal component  $S_t$  that is defined as the intra-year variation that is repeated constantly in an evolving fashion from year to year. The second component is the trend cycle component  $C_t$  that includes the variation due to the long-term trend, the business cycle, and other long-term cyclical factors. The third component  $D_t$  is the trading-day component that is comprised of the variation that can be attributed to the composition of the calendar. The fourth component making up the original series is the irregular component  $I_t$  (Lee *et al.*, 2000:829).

Since many economic time series, such as the maize data used in this project, are related in a multiplicative fashion ( $O_t = S_t C_t D_t I_t$ ) X11 compiles the seasonally-adjusted time series  $C_t I_t$  from the trend cycle and irregular components. In this multiplicative model, the trend cycle component  $C_t$  keeps the same scale as the original series  $O_t$ , while the

other three components vary around 1. The latter components are all expressed as percentages and therefore vary around 100.

The Census 12 procedure can be performed using a variety of software packages including SAS<sup>®</sup> (SAS, 2007a) and EViews' PROC X12 procedure (QMS, 2004b). Both these software packages were used to perform the X12 procedure with the X11 application for the various agricultural commodity series. For the purposes of this project, tests were only performed to find a seasonal factors series, and a trend-cycle (Henderson-curve) series. The seasonal factor series was utilised to model the seasonal patterns in the commodity data, while both the seasonal factors series and the Henderson-curve were used during the pairs selection process described in chapter 6.<sup>49</sup> The seasonal factors series and the Henderson-curve were utilised as new, additional pairs selection tools, adding to the existing pairs selection tools discussed in chapter 6. The X12 procedure was used to seasonally adjust the various agricultural commodity series by invoking the X11-ARIMA statement.

To seasonally adjust time series data, symmetric moving averages are used in estimating the various components. At the end of the series, however, these symmetric weights cannot be applied because asymmetric weights have to be used to extend the series. While various methods of extending a series have been proposed, the most noteworthy method to date has been the X11-ARIMA method developed at Statistics Canada (SAS, 2006c).<sup>50</sup>

The X11-ARIMA method makes use of an ARIMA model to forecast the series one or more years ahead in time. These forecasts are then appended to the original data series in order to improve the estimation of the seasonal factors, and so reduce revisions to the seasonally-adjusted series as new data becomes available (SAS, 2006c).

---

<sup>49</sup> Seasonal factors are events that cause normal fluctuations in business activity within individual or combinations of industries, while the Henderson-curve comprises of the long-run trend plus the business cycle (Lee *et al.* 2000:829)

<sup>50</sup> See section 4.3 for a discussion on ARIMA modelling.

## 5.4.2 Methodology

### 5.4.2.1 Introduction

In order to get a general idea of the presence of seasonality in the maize data, tests were only performed on the spot data for both white and yellow maize. The same procedure was repeated for all of the data series in chapter 6 only if the initial results were satisfactory. It is important to note that both the Henderson-curve and the seasonal factors series of each data set were compared in order to find pairs.

Section 5.4.2.2 below continues with the general PROC X12 model, followed by the results for white maize in section 5.4.2.3 and yellow maize in section 5.4.2.4. After testing for seasonal patterns in both the white and yellow maize spot data, the Henderson-curve of the white maize data is compared to the Henderson-curve of the yellow maize data in section 5.4.2.5. Similarly, the seasonal factors series of the white maize data are compared to the seasonal factors series of the yellow maize data in the same section.

### 5.4.2.2 The General Model

Both SAS<sup>®</sup> and EViews' PROC X12 procedures allow for the automatic selection of the ARIMA model that will best model the data. This can be done by specifying the AUTOMDL statement in PROC X12 (SAS<sup>®</sup>). When AUTOMDL is specified, the X12 procedure compares a selected model making use of the TRAMO (Time series Regression with ARIMA noise, Missing values, and Outliers) method by Gomez and Maravall (1996) to a default model.

The TRAMO method was implemented first, and involved two parts: identifying the orders of differencing and identifying the ARIMA model. This ARIMA model was selected from a collection of the best five pre-selected models. The best five models were selected from a wider range of models. Only the five models with the lowest Bayesian Information Criterion 2 (BIC2) value were considered during the next stage, in which the final model was selected.

During the next step, the program compared the model selected by the TRAMO method to a default model. The best model was then selected, which made it possible for the default model to be selected over the TRAMO model (SAS, 2006c). In deciding on the correct ARIMA model, three criteria were checked: the Mean Absolute Percent Error (MAPE) for the last three years of the series, the significance probability for the Box-Ljung Chi-square for twenty-four monthly lags, and over-differencing. A model that failed any one of these three criteria was excluded from further consideration.

The MAPE criterion is computed as follows (SAS, 2006c):

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|}. \quad (5.14)$$

For the model to meet the criteria, the MAPE for the last three years of the series had to be less than 15 percent.

The second criterion tests the lack of fit by making use of the model residuals. This was done by making use of the Box-Ljung formula (Ljung & Box, 1978:298):

$$\chi_m^2 = n(n+2) \sum_{k=1}^m (n-k)^{-1} r_k^2, \quad (5.15)$$

where  $n$  is the number of residuals that can be computed for the time series, and

$$r_k = \frac{\sum_{t=1}^{n-k} a_t a_{t+k}}{\sum_{t=1}^n a_t^2}, \quad (5.16)$$

where the  $a_t$ 's are the residual sequence. For the model to meet the second criterion, the Box-Ljung Chi-square for twenty-four lags had to be greater than 0.05 (SAS, 2006c).

The third criterion states that the over-differencing criteria must not exceed 0.9, and the model was thus not accepted if the over-differencing criteria exceeded this figure. The final set of all the models considered was those that passed all three criteria and that were estimated successfully. From this final set, the model with the smallest MAPE for the last three years was chosen. In the case where all five models failed, ARIMA

processing was skipped, and the standard X11 seasonal adjustment was performed (SAS, 2006c).

Once the correct ARIMA model was specified, a seasonally-adjusted series could be constructed. This was done by making use of the multiplicative adjustment option (default option in X11). In using multiplicative adjustment, the original series  $O_t$  is assumed to be of the form:

$$O_t = C_t S_t I_t P_t D_t, \quad (5.17)$$

where  $C_t$  is the trend cycle component,  $S_t$  is the seasonal component,  $I_t$  is the irregular component,  $P_t$  is the prior monthly factors component, and  $D_t$  is the trading-day component.

The standard X11–ARIMA seasonal adjustment method consists of four main steps. These steps are applied to the original data or the original data extended by an ARIMA model (SAS, 2006c). In the first step, the data are read, ignoring missing values until the first non-missing value is found. If prior monthly factors are present, the procedure reads the  $P_t$  (prior monthly) factors and divides them into the original series to obtain (SAS, 2006c):

$$\frac{O_t}{P_t} = C_t S_t I_t D_{tr,t} D_{r,t}. \quad (5.18)$$

The monthly factors developed in this process (or prior monthly factors) are then divided into the original or prior adjusted series to obtain:

$$C_t S_t I_t D_{r,t}. \quad (5.19)$$

In steps 2, 3 and 4, three iterations are performed, each of which provides estimates of the seasonal ( $S_t$ ), trading-day ( $D_{r,t}$ ), trend cycle ( $C_t$ ), and irregular components ( $I_t$ ). The estimates of the extreme values in the irregular components are refined

throughout all the iterations. After extreme values are identified and modified, final estimates of all the components are produced (SAS, 2006c).

Step 2 consists of three sub-steps: (a) During the first iteration in step 2, a centred, twelve-term moving average is applied to the original series  $O_t$  in order to provide a preliminary estimate  $\hat{C}_t$  of the trend cycle curve  $C_t$ . This moving average combines thirteen (a two-term moving average of a twelve-term moving average) consecutive monthly values, removing the  $S_t$  and  $I_t$ . Next, it obtains a preliminary estimate  $\widehat{S_t I_t}$ :

$$\widehat{S_t I_t} = \frac{O_t}{\hat{C}_t}. \quad (5.20)$$

(b) During the second step the estimated  $\hat{S}_t$ , of the seasonal factors is obtained by applying a moving average to the  $\widehat{S_t I_t}$ , which is then divided by  $\hat{S}_t$  to obtain an estimate  $\hat{I}_t$  of the irregular component. Next, a moving standard deviation is calculated from the irregular component and is used in assigning a weight to each monthly value for measuring its degree of extremeness (SAS, 2006c). These weights are used to modify extreme values in  $\widehat{S_t I_t}$ . A moving average is applied to the modified value of  $\widehat{S_t I_t}$  in order to estimate new seasonal factors. The original series is divided by these new seasonal factors to obtain a preliminary seasonally-adjusted series. Thereafter, a second estimate of the trend cycle is obtained by applying a weighted moving average to this seasonally-adjusted series (SAS, 2006c).

(c) During sub-step 3, the same computations are used to perform a second iteration on the original series. The same process is used to obtain second estimates of the seasonally-adjusted series and improved estimates of the irregular component. This irregular component is again modified for extreme values.  $\hat{I}_t$  is then used to provide estimates of trading-day factors, and as refined weights for the identification of extreme values (SAS, 2006).

During step 3, this series is adjusted by the trading-day factors and irregular weights that are developed during the first iteration. The second iteration produces final estimates of the trading-day factors and irregular weights (SAS, 2006c).

During step 4, a third and final iteration is performed by making use the original series that has been adjusted for trading-day factors, as well as the irregular weights that were computed during the second iteration. During the third iteration, PROC X12 develops final estimates of the seasonal factors series, the seasonally-adjusted series, the trend cycle series, and the irregular components. The procedure then computes summary measures of variation and produces a moving average of the final adjusted series (Lee *et al.*, 2000:832).

All of these steps were repeated for each of the commodity series. Section 5.4.2.3 below reports on the test for seasonality in the white maize data that made use of the PROC X12 and X11–ARIMA procedures. Section 5.4.2.4 reports on the same procedures for yellow maize data. In section 5.4.2.5, the results of the white and yellow maize tests are compared. Please note that the data series used for these tests are identical to those series used in testing for seasonality with GARCH.<sup>51</sup>

As with the GARCH tests, tests for seasonality were first performed on the WMSPOT and YMSPOT series. If the X12 procedure is found to be successful in identifying the seasonal patterns in the WMSPOT and YMSPOT series, all of the other series will also be tested by means of the same process.

### 5.4.2.3 White Maize

#### 5.4.2.3.1 Model Results

The ARIMA model allows for various outputs that are useful for the recognition of seasonal patterns. The first is a seasonally-adjusted series that is obtained by the process explained in the steps in section 5.4.2.1 above. The second is a seasonal factors series, and the third is the Henderson-curve (QMS, 2004a:326).<sup>52</sup> The final model is derived

---

<sup>51</sup> See section 5.2 above for a description of the data.

<sup>52</sup> Only the Henderson-curve and the seasonal factors series are important for the purpose of this thesis.

from the more general seasonal ARIMA (SARIMA) model. This general model can be given as (Chatfield, 2004:66):

$$\Phi_p(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S)W_t, \quad (5.21)$$

where  $B$  is the backshift operator;  $\phi_p$ ,  $\Phi_p$ ,  $\theta_q$ ,  $\Theta_Q$  are polynomials of order  $p$ ,  $P$ ,  $q$ , and  $Q$  respectively, and  $W_t$  follows a random process. The above model is thus of the order  $(p, d, q) (P, D, Q)$ , where  $p$  is the non-seasonal AR order,  $d$  the order of non-seasonal differences,  $q$  the non-seasonal MA order,  $P$  the seasonal AR order,  $D$  the order of seasonal differences, and  $Q$  the seasonal MA order.

In order to move from the SARIMA model to the more specific model used to fit the data, the first step is to assess the values of  $d$  and  $D$  which removes most of the trend and seasonality from the data. Thereafter, the values of  $p$ ,  $P$ ,  $q$  and  $Q$  are usually assessed by looking at the autocorrelation and partial autocorrelation functions of the differenced series (Chatfield, 2004:67). For the purpose of finding the Henderson-curve and the seasonal factor series, the model used to extend the original series was selected by making use of the X12 automatic model selection process. The AUTOMDL statement identified the following five models.

TABLE 5.7 BEST FIVE MODELS CHOSEN BY THE AUTOMATIC MODELLING PROCESS

Estimated Model							
Rank	$p$	$d$	$q$	$P$	$D$	$Q$	BIC2
1	1	1	0	0	0	0	-1.882
2	0	1	1	0	0	0	-1.872
3	1	1	0	0	0	1	-1.842
4	1	1	0	1	0	0	-1.842
5	1	1	1	0	0	0	-1.841

During this process the five pre-selected models were tested against the three criteria discussed in section 5.4.2.2 above. The MAPE value of the final model was 2.58

percent over the last three years and falls within the range of 15 percent.<sup>53</sup> The model selected by the AUTOMDL option in PROC X12 for the white maize series was an ARIMA model of the order (1,1,0) (0,0,0) and can be given by:

$$Z_t = (1 + \phi)Z_{t-1} - \phi Z_{t-2} + W_t. \quad (5.22)$$

This model is a non-seasonal model in that the seasonal orders ( $P$ ,  $D$ ,  $Q$ ) are all zero, indicating that the model fits a data series without any definite seasonal pattern.

To test for seasonality, three tests were performed. The first is a test for 'stable seasonality' and is a one-way analysis of the variance on the final unmodified SI-ratios using seasons as the factor. A large F-value and small significance level is evidence that a significant amount of variation in the SI-ratios is due to month effects. In turn, this variation in the SI-ratios is evidence of seasonality, leading to a rejection of the null hypothesis of no month effect (SAS, 2006).

The second test is an F-test for moving seasonality and is performed by a two-way analysis of variance. The two factors used are seasons (months or quarters) and years. Tests for the year effect are performed separately, where the null hypothesis is that there is no effect due to year effects, after accounting for variation due to months or quarters. Please note that the significance level reported in both the moving and stable seasonality tests is only an approximate value.

The third test is a Kruskal-Wallis Chi-squared test for stable seasonality. The test for identifiable seasonality was thus performed by combining the F-tests for stable and moving seasonality, along with the Kruskal-Wallis test for stable seasonality. This combined test was performed according to the process detailed below (Dagum, 1980). Tests for identifiable seasonality were performed first. If the null hypothesis in the moving seasonality tests is not rejected at the 0.10 percent, the seasonality is not identifiable. If the null hypothesis in step one is rejected, but the moving seasonality null hypothesis is not rejected at the 5 percent level, the following quantities are computed:

---

<sup>53</sup> Please note that SAS® do not report the Ljung-Box statistic and over-differencing values. Only the MAPE value for the last three years is specified.

$$T_1 = \frac{7}{F_m - F_s}, \text{ and } T_2 = \frac{3F_m}{F_s}. \quad (5.23)$$

If  $T$  denotes the simple average of  $T_1$  and  $T_2$ :

$$T = \frac{(T_1 + T_2)}{2}, \quad (5.24)$$

then the null hypothesis of identifiable seasonality not present is accepted if  $T \geq 1$ .

If the moving seasonality F-test based on  $F_m$  passes, but one of the two statistics based on the  $T$  quantities fails, or the Kruskal-Wallis Chi-squared test fails at the 1 percent level, then there is no identifiable seasonality present in the data (the null hypothesis is not rejected). However, if the  $F_s$ ,  $F_m$  and Kruskal-Wallis Chi-squared tests pass, the null hypothesis of identifiable seasonality not present is rejected.

The tests performed on the white maize spot data, indicated that there is no identifiable seasonality present. The F-value of the first test (stable seasonality) is 4.33, with a significance value of less than 0.1 percent, thus indicating that stable seasonality is present in the white maize data. The Kruskal-Wallis Chi-squared test statistic for stable seasonality is 35.87 with a probability level of 0.02 percent, also indicating the presence of stable seasonality. The F-test value for moving seasonality is 1.40, and this test indicated that no moving seasonality is present at the 5 percent level of statistical significance.

Even though no moving seasonality is present in the white maize data, the seasonal factors series and the Henderson-curve could still prove useful as pairs selection tool. The seasonal factors and Henderson-curve-series for the white maize data are discussed in section 5.4.2.5 where they are compared to the seasonal factors and Henderson-curve-series for the yellow maize data.

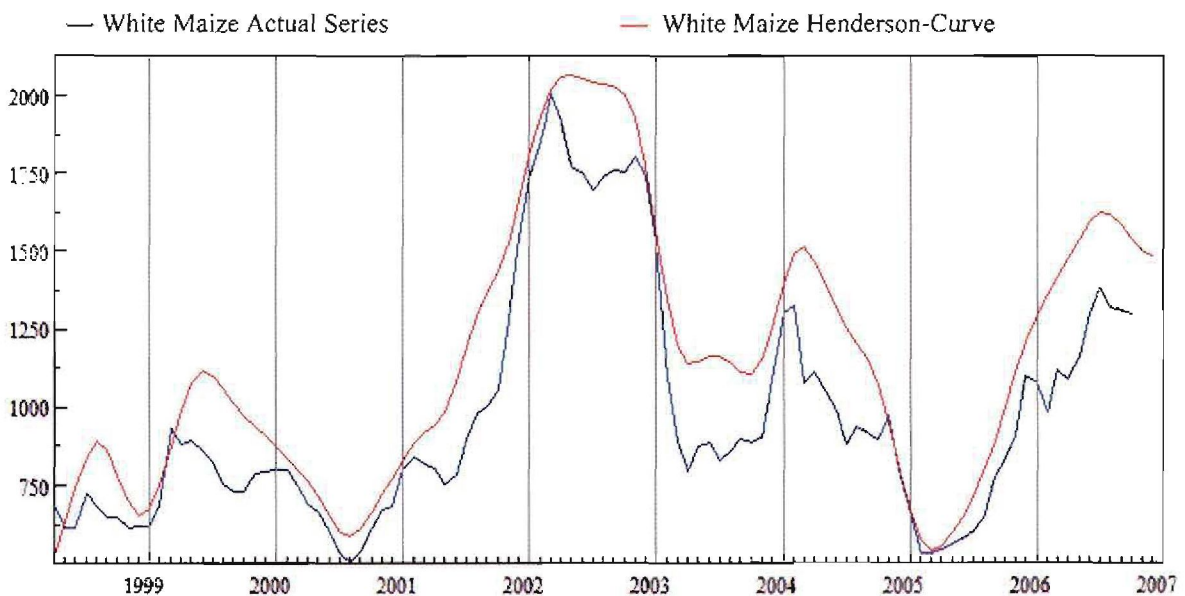
#### 5.4.2.3.2 Final Results – White Maize

Even though PROC X12 is traditionally used in conjunction with X11-ARIMA as a tool for seasonal adjustment, this process was utilised for the purpose of selecting pairs

from the South African maize data for this project. As can be seen in graph 5.1 below, the Henderson-curve produced during the PROC X12 process gives an accurate estimate of the original white maize series. The Henderson-curve can be used in two ways by the pairs trader: in predicting possible future scenarios regarding the value of his pairs portfolio, and by comparing the Henderson-curves of different series as a pairs selection tool.

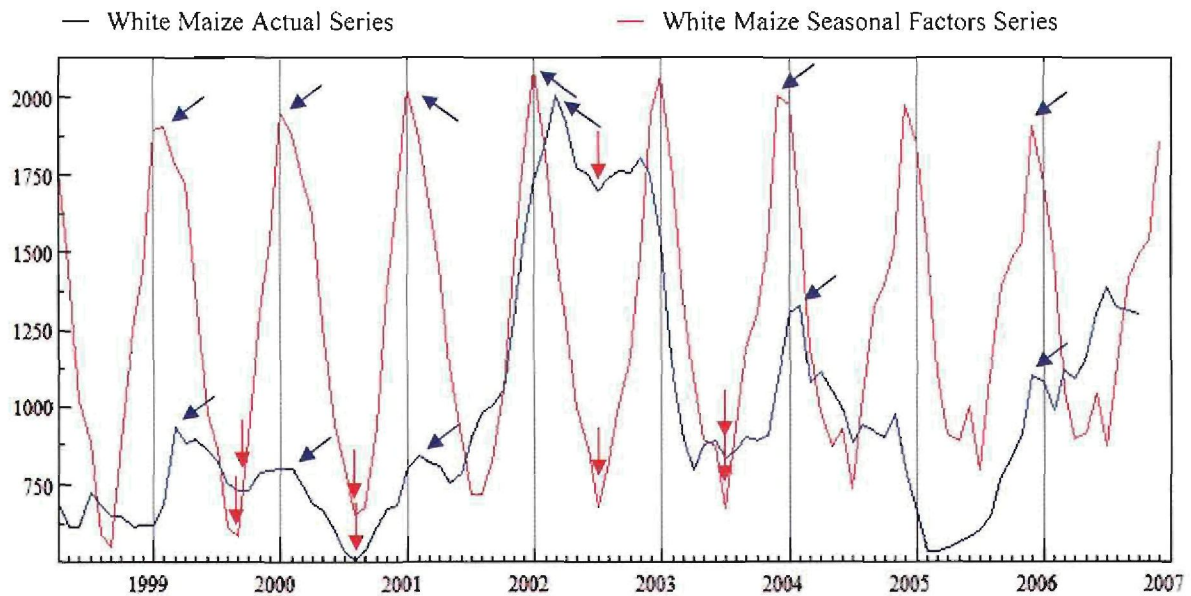
Graph 5.2 illustrates the relationship between the actual series and seasonal factors series.<sup>54</sup> This comparison shows that the seasonal factors series did well to explain the movements in the actual data for the complete period of the data. This series could be useful for the pairs trader as an indicator regarding future price movements. As with the Henderson-curve, the seasonal factors series were used as a pairs selection tool.

GRAPH 5.1 MONTHLY AVERAGE SPOT PRICE SERIES OF WHITE MAIZE VS. HENDERSON-CURVE



<sup>54</sup> Please note that the prices of the two maize series are given on the y-axis while the date is shown on the x-axis for graphs 5.1 and 5.2.

GRAPH 5.2 MONTHLY AVERAGE SPOT PRICE SERIES OF WHITE MAIZE VS. SEASONAL FACTORS SERIES



The blue arrows in graph 5.2 above signify the peaks of both the seasonal factors series and the actual spot series where the seasonal factor series leads the actual spot series. The red arrows show the cases where the seasonal factor series leads the actual series for the troughs in each cycle. Section 5.4.2.4 below continues with the PROC X12 tests for the yellow maize spot data.

#### 5.4.2.4 Yellow Maize

##### 5.4.2.4.1 Model Results

As with the white maize series, the seasonally-adjusted series for the yellow maize data was obtained by the process explained in the steps in section 5.4.2.2 above. In finding the Henderson-curve and the seasonal factor series in the yellow maize data, the model used to extend the original data series, was selected by making use of the X12 process. The AUTOMDL statement identified the following five models.

TABLE 5.8 BEST FIVE MODELS CHOSEN BY THE AUTOMATIC MODELLING PROCESS

Estimated Model							
Rank	$p$	$d$	$q$	$P$	$D$	$Q$	BIC2
1	1	1	0	0	0	0	-2.418
2	0	1	1	0	0	0	-2.387
3	2	1	0	0	0	1	-2.377
4	1	1	1	1	0	0	-2.377
5	1	1	0	0	0	1	-2.376

The MAPE value of the final model was 2.37 percent over the last three years, and falls within the range of 15 percent. The model selected by the AUTOMDL option in SAS<sup>®</sup>'s PROC X12 for the yellow maize series was an ARIMA model of the order (1,1,0) (0,0,0) and is given by:

$$Z_t = (1 + \phi)Z_{t-1} - \phi Z_{t-2} + W_t. \quad (5.25)$$

This model is a non-seasonal model in that the seasonal orders ( $P$ ,  $D$ ,  $Q$ ) are all zero, indicating that the model fits a data series without any definite seasonal pattern. As with the white maize data three tests were performed to test for identifiable seasonality.

The first test for 'stable seasonality' performed on the yellow maize spot data, indicated that stable seasonality is present in the yellow maize spot data. The F-value for the first test (stable seasonality) is 11.92 and has a significance value of less than 0.1 percent, evidence that stable seasonality is present in the yellow maize data.

The Kruskal-Wallis Chi-squared test statistic for stable seasonality is 65.42 with a probability level of 0.00 percent, supporting the evidence of the presence of stable seasonality. The F-test (2.33) for moving seasonality indicated that moving seasonality is present at the 5 percent level of statistical significance.

The final result of the three tests is that no identifiable seasonality is present in the yellow maize data. Even though no moving seasonality is present in the yellow maize

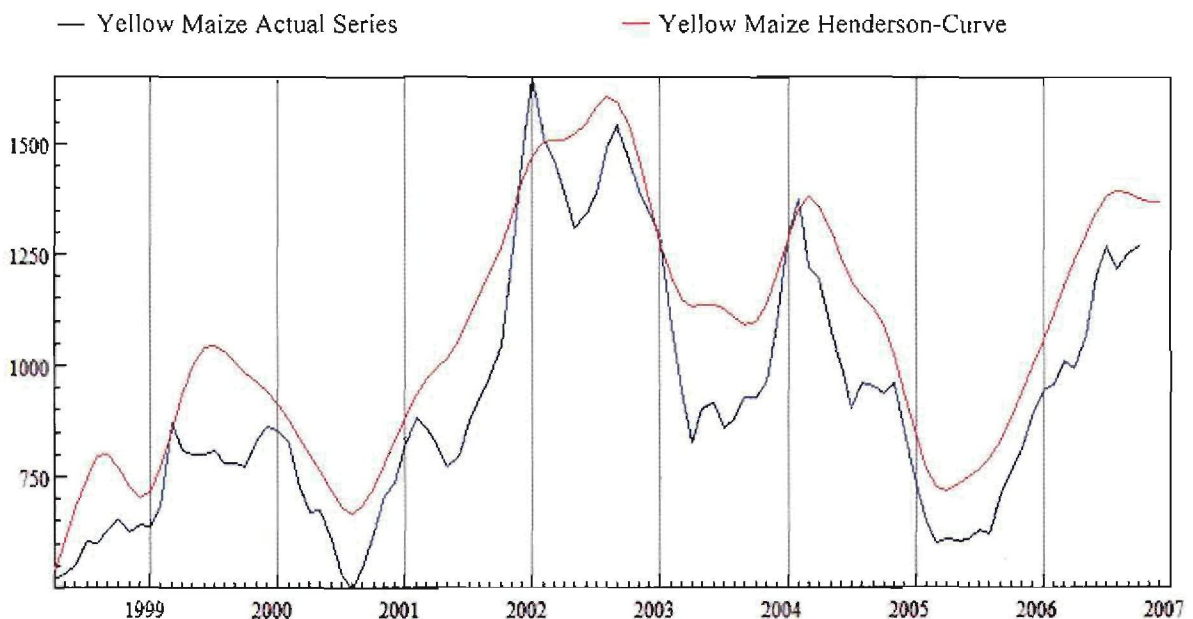
data, the Henderson-curve and the seasonal factors series were still utilised as pairs selection tools in chapter 6.

#### 5.4.2.4.2 Final Results – Yellow Maize

As can be seen in graph 5.3 below, the Henderson-curve produced during the PROC X12 process gives an accurate estimate of the original yellow maize series. As with the white maize data, the Henderson-curve was used in predicting possible future scenarios regarding the value of a pairs portfolio, as well as an additional tool used for pairs selection.<sup>55</sup>

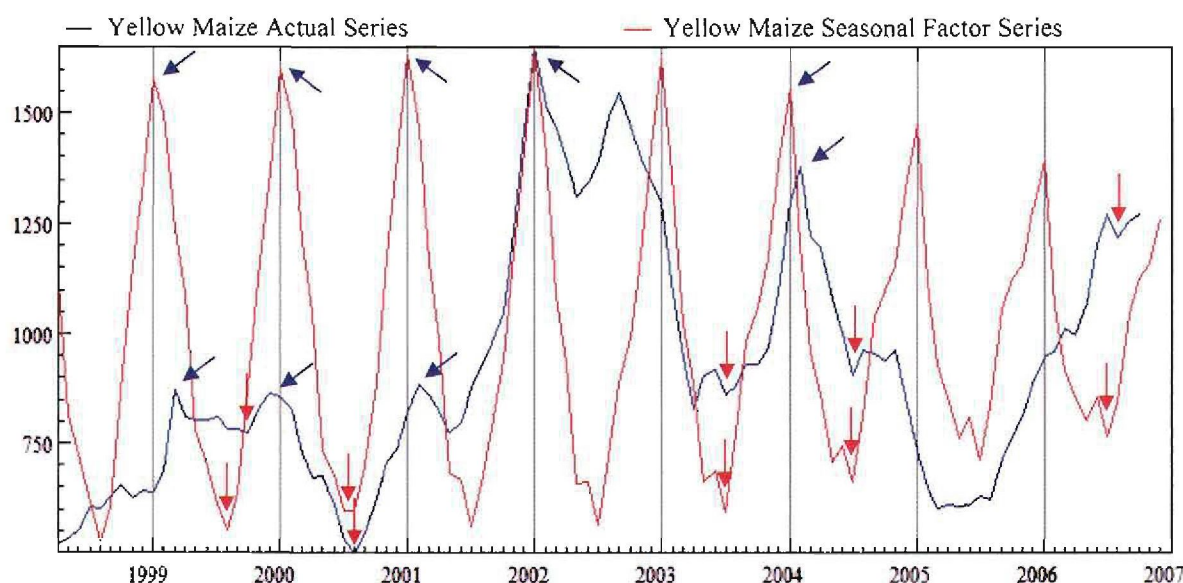
In graph 5.4 the relationship between the actual yellow maize spot series and seasonal factors series are given. This comparison shows that the seasonal factors series did equally well. The seasonal factors series for the yellow maize data was also used for predicting possible future scenarios regarding the value of a pairs portfolio and as pairs selection tool.

GRAPH 5.3 MONTHLY AVERAGE SPOT PRICE SERIES OF YELLOW MAIZE VS HENDERSON-CURVE



<sup>55</sup> Pairs will be selected based on their long-run relationship as indicated by the correlation between their Henderson-curves.

GRAPH 5.4 MONTHLY AVERAGE SPOT PRICE SERIES OF YELLOW MAIZE VS. SEASONAL FACTORS SERIES



The blue arrows in graph 5.4 above signify the instances where the seasonal factor series peaks before the actual spot series. The red arrows indicate the cases where the seasonal factor series reaches its low ahead of the actual series. As was the case with the white maize data, both the Henderson-curve and the seasonal factors series did well in following the actual series and can be utilised as an indication of the general direction the market is moving in.

Section 5.4.2.5 below continues with a comparison between the Henderson-curves and seasonal factors series of the white and yellow maize spot series. It is important to note that this comparison between the two spot series only serve as an example of what was done for all the different pairs given in chapter 6.

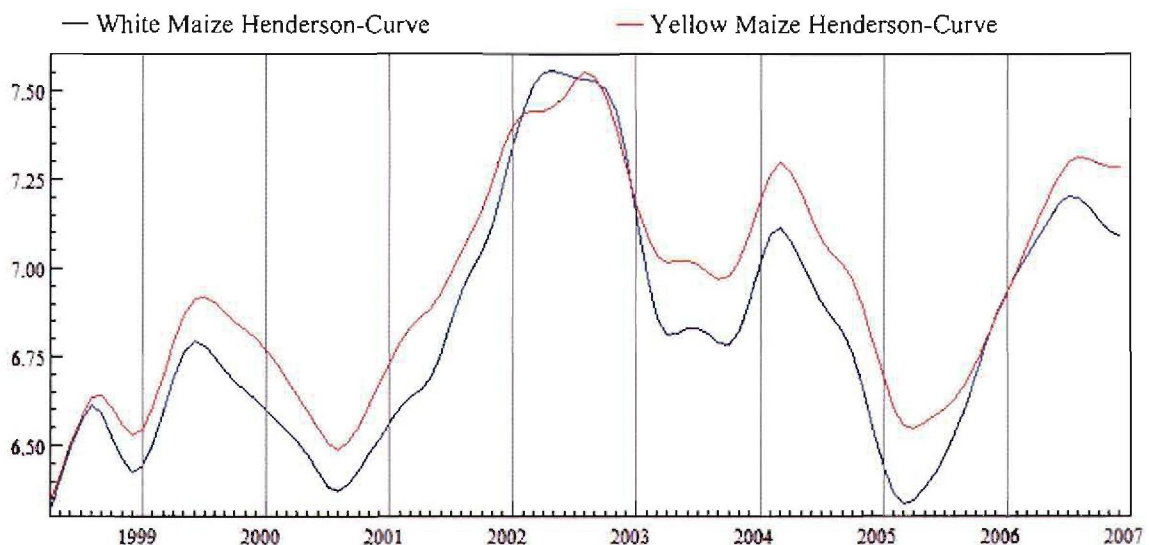
#### 5.4.2.5 White Maize vs. Yellow Maize

As mentioned above, ARIMA can be utilised in various ways. Firstly, this deterministic method for testing seasonality may be used by the commodity pairs trader as guide regarding the seasonal behaviour of the underlying commodities that he is trading. Secondly, when considering both the Henderson-curve and seasonal factors results of the two spot series, it is evident that they may be useful as additional tools for selecting pairs.

To test the Henderson-curve as pairs selection tool, the Henderson-curve of the white maize spot series would have to be compared to that of the yellow maize spot series. If the two Henderson-curves match each other closely, they can be considered a good candidate for a pair.<sup>56</sup> Graph 5.5 below compares the Henderson-curves of the white and yellow maize spot series as an example of how these curves were used as a pairs selection tool as will be detailed in chapter 6.<sup>57</sup>

The test for the seasonal factors series as pairs selection tool was done in the same way as that of the Henderson-curve. If the two seasonal factors series match each other closely, they can be considered a good candidate for a pair. Graph 5.6 below compares the seasonal factors series of the white and yellow maize spot series as an example of how these series were used for the different combinations of pairs as a pairs selection tool as will be discussed in chapter 6.<sup>58</sup>

GRAPH 5.5 WHITE MAIZE HENDERSON-CURVE VS. YELLOW MAIZE HENDERSON-CURVE

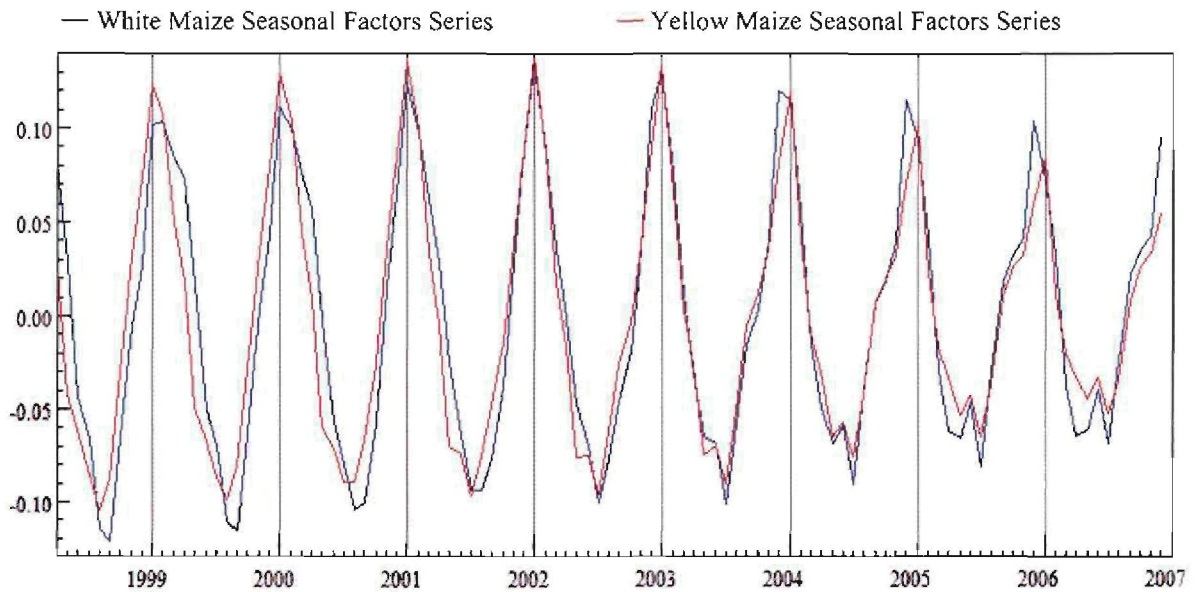


<sup>56</sup> Please note that various other checks need to be performed before two series can be considered a pair, see Chapter 6.

<sup>57</sup> Graph 5.5 only serves as an example of the comparisons that were performed, as reported on in chapter 6, for all the possible pairs combinations. X12 was only used as an additional pairs selection tool, in conjunction with the ARIMA tests as well as the traditional pairs selection tools that will be discussed in chapter 7. Note that the Henderson-curve data is given on the y-axis, while the date is given on the x-axis.

<sup>58</sup> Note that the seasonal factors data is given on the y-axis, while the date is given on the x-axis.

GRAPH 5.6 WHITE MAIZE SEASONAL FACTORS SERIES VS. YELLOW MAIZE SEASONAL FACTORS SERIES



The Henderson-curve and the seasonal factors series can also be used as a tool to identify entry and exit points once pairs have been identified. In order to apply these series in such a manner, one of the price series was subtracted from the other in order to determine where the two series are the furthest apart. If, for example, the difference between the white maize spot series' Henderson-curve and that of the yellow maize spot series' starts to decrease after a reaching a high, the pairs trader will identify this as a signal to buy the yellow maize spot series whilst simultaneously short selling the white maize spot series.

The same procedure was used in identifying buying and selling signals with the aid of the seasonal factors series. However, it is to be noted that these two series (Henderson-curve and seasonal factors series) were used in conjunction with fundamental analysis as well as some technical analysis tools.

Since the X12 process was successful in identifying the seasonal patterns in both the white and yellow maize spot series, these tests were repeated for all the maize series. Section 5.4.2.6 reports on the results of all the tests for seasonality. The Henderson-curve and seasonal factors series extracted from the various maize series were compared with one another during the pairs selection process described in chapter 6.

#### 5.4.2.6 Testing for Seasonality in the South African Maize Data with X12

In sections 5.4.2.3 and 5.4.2.4 above, X12 was used to test for seasonal patterns in the white and yellow maize spot series data. These tests were repeated for all of the maize series with the aim of extracting the Henderson-curve and the seasonal factors series of each maize series. These series will be compared to each other in chapter 6 with the aim of selecting pairs, and are thus not given in this chapter.

The first step in the estimation process is to identify five models best fitting the data for every one of the eight different maize data series. The best five models are reported in Table 5.9 below.<sup>59</sup>

---

<sup>59</sup> Please refer to section 5.4.2.2 for an explanation regarding the selection of the correct ARIMA model.

TABLE 5.9 BEST FIVE PREDEFINED MODELS

Agricultural Commodity Series	Estimated Model							
	Rank	$p$	$d$	$q$	$P$	$D$	$Q$	BIC2
WM SPOT	1	1	1	0	0	0	0	-1.882
	2	0	1	1	0	0	0	-1.872
	3	1	1	0	0	0	1	-1.842
	4	1	1	0	1	0	0	-1.842
	5	1	1	1	0	0	0	-1.841
WM 1 away	1	1	1	0	0	0	0	-1.947
	2	0	1	1	0	0	0	-1.947
	3	1	1	0	0	0	1	-1.909
	4	1	1	0	1	0	0	-1.909
	5	1	1	1	0	0	0	-1.908
WM 2 away	1	0	1	1	0	0	0	-1.970
	2	1	1	0	0	0	0	-1.963
	3	0	1	1	0	0	1	-1.940
	4	0	1	1	1	0	0	-1.940
	5	1	1	1	0	0	0	-1.928
WM 3 away	1	0	1	1	0	0	0	-1.984
	2	1	1	0	0	0	0	-1.977
	3	0	1	1	1	0	0	-1.943
	4	0	1	1	0	0	1	-1.943
	5	0	1	2	0	0	0	-1.942
YM SPOT	1	1	1	0	0	0	0	-2.418
	2	0	1	1	0	0	0	-2.387
	3	2	1	0	0	0	0	-2.377
	4	1	1	1	0	0	0	-2.377
	5	1	1	0	0	0	1	-2.376
YM 1 away	1	0	1	1	0	0	0	-4.218
	2	1	1	0	0	0	0	-4.200
	3	1	1	1	0	0	0	-4.175
	4	0	1	1	0	0	1	-4.174
	5	0	1	1	1	0	0	-4.174
YM 2 away	1	0	1	1	0	0	0	-2.589
	2	0	1	1	0	0	1	-2.553
	3	0	1	1	1	0	0	-2.552
	4	1	1	0	0	0	0	-2.545
	5	0	1	2	0	0	0	-2.545
YM 3 away	1	0	1	1	0	0	0	-2.568
	2	0	1	1	0	0	1	-2.527
	3	0	1	1	1	0	0	-2.527
	4	1	1	1	0	0	0	-2.525
	5	0	1	2	0	0	0	-2.525

From these models (the best five for every maize series), the TRAMO process selected the best models that were then used to extend the original series. As mentioned in section 5.4.2.2, the correct ARIMA model was chosen based on the following three criteria: the Mean Absolute Percent Error (MAPE) for the last three years of the series, the significance probability for the Box-Ljung Chi-square for twenty-four monthly lags, and over-differencing. The MAPE values and the order of each of the various final models are given by Table 5.10 below.<sup>60</sup>

TABLE 5.10 BEST FIVE PREDEFINED MODELS

Agricultural Commodity Series	WM Spot	WM 1 away	WM 2 away	WM 3 away	YM Spot	YM 1 away	YM 2 away	YM 3 away
MAPE	2.58	2.47	2.15	1.92	2.37	2.25	2.01	2.44
Model order	(1,1,0) (0,0,0)	(1,1,0) (0,0,0)	(0,1,1) (0,0,0)	(0,1,1) (0,0,0)	(1,1,0) (0,0,0)	(0,1,1) (0,0,0)	(0,1,1) (0,0,0)	(0,1,1) (0,1,1)

The final models selected for each of the maize series are given by the following equations 5.24 to 5.26. The WMSPOT, WM 1 away and YMSPOT series take the form of:

$$Z_t = (1 + \phi)Z_{t-1} - \phi Z_{t-2} + W_t. \quad (5.26)$$

WM 2 away, WM 3 away, YM 1 away and YM 2 away take the form of:

$$Z_t = Z_{t-1} + W_t - \theta W_{t-1}, \quad (5.27)$$

while YM 3 away takes the form of:

$$Z_t = Z_{t-1} + Z_{t-12} - Z_{t-13} - \theta W_{t-1} - \theta W_{t-12} + \theta \theta W_{t-13}. \quad (5.28)$$

The tests for identifiable seasonality revealed that no identifiable seasonality is present for WMSPOT, WM 1 away, WM 2 away, WM 3 away, YM 2 away and YM 3 away. However, identifiable seasonality is present for YMSPOT and YM 1 away. Table 5.11 below details the results of the seasonality tests.<sup>61</sup>

<sup>60</sup> Note that all of the model orders are given in the form  $(p,d,q)(P,D,Q)$ .

<sup>61</sup> Please refer to section 5.4.2.3.1 above for an explanation of the tests for seasonality.

TABLE 5.11 TESTS FOR IDENTIFIABLE SEASONALITY

Agricultural Commodity Series	Tests for seasonality	Statistic value	<i>p-value</i>	Presence of seasonality
WM Spot	F-value assuming stability	4.325	< 0.1 %	Identifiable seasonality probably not present
	Kruskal-Wallis Statistic	35.871	0%	
	F-value moving seasonality	1.403	> 5 %	
WM 1 away	F-value assuming stability	3.658	< 0.1 %	Identifiable seasonality not present
	Kruskal-Wallis Statistic	31.466	0%	
	F-value moving seasonality	2.227	> 5 %	
WM 2 away	F-value assuming stability	1.308	< 0.1 %	Identifiable seasonality not present
	Kruskal-Wallis Statistic	13.893	24%	
	F-value moving seasonality	2.006	> 5 %	
WM 3 away	F-value assuming stability	1.567	< 0.1 %	Identifiable seasonality not present
	Kruskal-Wallis Statistic	16.622	12%	
	F-value moving seasonality	1.746	> 5 %	
YM Spot	F-value assuming stability	11.917	< 0.1 %	Identifiable seasonality present
	Kruskal-Wallis Statistic	65.417	0.00%	
	F-value moving seasonality	2.332	> 5 %	
YM 1 away	F-value assuming stability	8.559	< 0.1 %	Identifiable seasonality present
	Kruskal-Wallis Statistic	53.859	0.00%	
	F-value moving seasonality	1.734	> 5 %	
YM 2 away	F-value assuming stability	2.023	< 0.1 %	Identifiable seasonality not present
	Kruskal-Wallis Statistic	21.697	2.70%	
	F-value moving seasonality	0.894	> 5 %	
YM 3 away	F-value assuming stability	2.263	< 0.1 %	Identifiable seasonality not present
	Kruskal-Wallis Statistic	22.358	2.20%	
	F-value moving seasonality	0.981	> 5 %	

Even though identifiable seasonality is not present in the majority of the maize series, the underlying seasonal patterns in the data are well forecasted by the Henderson-curve and the seasonal factors series of each of the maize data series.

### 5.4.3 Conclusion

In testing for seasonality with X12, better results were achieved than with the tests performed with GARCH. The seasonal factors series acts as a leading indicator for the white and yellow maize series for a large portion of the test period. Because both the seasonal factors series and the Henderson-curve did well to identify the seasonal

patterns in the original data series, X12 was extended to the other data series to test for pairs as reported in chapter 6.

X12 can thus be employed for two purposes by the pairs trader: as a tool to identify seasonal patterns in the maize data and as pairs selection tool. Section 5.5 discusses the testing for seasonal patterns by making use of unobserved components modelling.

## 5.5 Modelling for Seasonal Patterns with UCM

### 5.5.1 Introduction

As previously stated, Unobserved Component Modelling (UCM) is a stochastic approach to modelling time series data. Although UCM is not frequently used in the literature (Harvey & Scott, 1994; Al-Saad & Moosa, 2005; and Al-Deehani, 2006). Section 5.5.2 presents the methodology discussed in the literature regarding the use of UCM to test for hidden seasonality in the agricultural commodities data. As mentioned previously, UCM was employed in this project in two ways: as a tool to identify hidden seasonal patterns in the data, and as an additional pairs selection tool.

### 5.5.2 Methodology

#### 5.5.2.1 Estimating the Model – General Model Specification

In testing for hidden monthly seasonal effects, it is necessary to specify the correct model. This was done in terms of four components: the trend  $\mu_t$ , the cyclical component  $\varphi_t$ , the seasonal component  $\gamma_t$ , and the random component  $\varepsilon_t$ . The same tests were performed in both PROC UCM in SAS<sup>®</sup> (2007a) and the Structural Time Series Analyser, Modeller and Predictor (STAMP<sup>™</sup>) package (OxMetrics, 2006). Several preliminary models were employed in order to decide on the final model. The process of determining the final model is discussed in sections 5.5.2.1.1 to 5.5.2.1.3 that follow.<sup>62</sup>

---

<sup>62</sup> Although several preliminary tests were performed, only the final model is discussed in detail here.

### 5.5.2.1.1 Modelling the Trend Component

The trend component consists of a level (the actual value of the trend) and a slope (the drift factor). The default setting in STAMP™ allows for both the level and slope to be stochastic. It is however possible to stipulate some special cases: random walk with noise (stochastic level and no slope), local level with drift (stochastic level and fixed slope), and a smooth trend (fixed level and a stochastic slope) (Koopman *et al.*, 2006:55). The same options are available when modelling the trend component in SAS®'s PROC UCM (2007a).

In order to assess which of the special cases were to be used for the purpose of testing the agricultural commodities data in the final model, the default setting was specified in the preliminary model to test for the statistical significance of the slope component.

### 5.5.2.1.2 Modelling the Cyclical Component

The parameters of the cycle are estimated from the data and consist of: a variance parameter  $\sigma_K^2$ , which is responsible for making the cycle stochastic; a period  $2\pi/\lambda$ ; a damping factor  $\rho$ ; a frequency (in radians)  $\lambda_c$ ; and a variance of the disturbance terms in its stochastic equations (SAS, 2007a). Contrary to a deterministic cycle where the sine-cosine wave has a given period, a stochastic cycle is constructed by shocking it with disturbances and introducing a damping factor (Koopman *et al.*, 2006:59).<sup>63</sup>

The statistical specification of the stochastic cycle is:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} v_t \\ v_t^* \end{bmatrix}, \quad t = 1, \dots, T. \quad (5.29)$$

In testing for seasonal effects in the agricultural commodities data a generalised specification of the cycle is used. This cycle is smoother and can be specified as  $\psi_t = \psi_t^{(k)}$  where:

<sup>63</sup> Please see section 4.5 for more on this.

$$\begin{bmatrix} \psi_t^{(j)} \\ \psi_t^{*(j)} \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1}^{(j)} \\ \psi_{t-1}^{*(j)} \end{bmatrix} + \begin{bmatrix} \psi_t^{(j-1)} \\ \psi_t^{*(j-1)} \end{bmatrix}, \quad (5.30)$$

for  $j=1, \dots, k$ ;  $t=1, \dots, n$ ; and where  $v_t = \psi_t^{(0)}$  and  $v_t^* = \psi_t^{*(0)}$  are mutually uncorrelated white noise disturbances with zero means and common variance  $\sigma_v^2$  (Koopman *et al.*, 2006:63). For this equation, the value of  $k$  determines the smoothness of the cycle.

### 5.5.2.1.3 Modelling the Seasonal Component

Seasonal effects are considered corrections to the general trend of the series due to the seasonal variations. These effects sum to zero over the full season cycle and are therefore modelled as a stochastic periodic pattern of an integer period  $s$ , such that the sum  $\sum_{i=0}^{s-1} \gamma_{t-i}$  is always zero in the mean.

As mentioned in section 4.5, the seasonal component can be modelled in either a dummy variable or trigonometric form. In SAS<sup>®</sup> (SAS, 2007a) the dummy variable form is the default. In order to run the test for the trigonometric form the *TYPE=TRIG* option should be specified in the PROC UCM module (SAS, 2007a). It is furthermore also possible to specify the length of the seasonal component. A seasonal component with season length four will correspond to quarterly seasonality while season length twelve indicates testing for monthly seasonality (SAS, 2007a).

All of these components were combined to test for any unobserved seasonal patterns in the agricultural commodities data. Section 5.5.2.2 below discusses the testing of unobserved seasonality in the white maize data, while section 5.5.2.3 discusses the testing for unobserved seasonality in the yellow maize data. Note that as before tests were first performed for the white and yellow maize spot series before repeating the same tests for the other maize series.

### 5.5.2.2 White Maize

#### 5.5.2.2.1 The Model

The final model used in testing for hidden seasonality in the white maize data took the form of:

$$y_t = \mu_t + \varphi_t + \gamma_t + \varepsilon_t. \quad (5.31)$$

where  $\mu_t$  is the trend,  $\varphi_t$  the cyclical component,  $\gamma_t$  the seasonal component, and  $\varepsilon_t$  the random component. During the preliminary testing phase the *p-value* for the slope component was reported at 0.2666 and that of the random component at 0.9996. These two components were therefore left out of the final model. For the final model, the trend component thus took the form of a random walk with noise. The model for  $\mu_t$  is thus given by:

$$\mu_t = \mu_{t-1} + \eta_t \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2), \quad (5.32)$$

where the disturbance term  $\eta_t$  is assumed to be independent. The trend component  $\mu_t$  is a random walk (the level) and is thus modelled as a time-varying level component without any persistent upward or downward drift (no slope component is included).

For the cyclical component Harvey and Trimbur (2003) were followed in making use of the smoothed cycle option in SAS<sup>®</sup> (SAS, 2007a). They prefer the smoothed cycle specification because of its better-fit characteristics. The final model used in testing for unobserved seasonality in the white maize data thus took the form of  $\psi_t = \psi_t^{(k)}$  where:

$$\begin{bmatrix} \psi_t^{(j)} \\ \psi_t^{*(j)} \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1}^{(j)} \\ \psi_{t-1}^{*(j)} \end{bmatrix} + \begin{bmatrix} \psi_t^{(j-1)} \\ \psi_t^{*(j-1)} \end{bmatrix}, \quad (5.33)$$

for  $j = 1, \dots, k$ ;  $t = 1, \dots, n$ ; and where  $v_t = \psi_t^{(0)}$  and  $v_t^* = \psi_t^{*(0)}$  are mutually uncorrelated white noise disturbances with zero means and common variance  $\sigma_v^2$ . In this case  $\rho < 1$ , and  $\psi_t$  thus has a stationary distribution with zero mean and variance  $\sigma_v^2 / (1 - \rho^2)$ . The value of  $k$  (the order of the cycle) determines the smoothness of the cycle.

For the seasonal component, Koopman *et al.* (2006:56) suggest that the seasonal component should rather be fixed and not allowed to be stochastic in the case where there is not enough data to allow a changing pattern to be estimated. Since there are only 103 months (data points) available, the seasonal component took the trigonometric form. In this form  $\gamma_t$  is modelled as a sum of cycles of different frequencies. This model is given as (Harvey & Scott, 1994:1328):

$$\sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t}, \quad (5.34)$$

where  $s$  equals  $(s-1)/2$ . The cycles  $\gamma_{j,t}$  have frequencies  $\lambda_j = 2\pi j/s$  and are specified by the matrix equation (SAS, 2006a:1633):

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \psi_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad (5.35)$$

where the disturbances  $\omega_{j,t}$  and  $\omega_{j,t}^*$  are assumed to be independent,  $j$  is fixed, and  $\omega_{j,t}$  and  $\omega_{j,t}^* \sim N(0, \sigma_\omega^2)$ . Since the seasonal component for this model is deterministic, the decomposition of the seasonal effects into these harmonics ( $\gamma_{j,t}$ ) is identical to its Fourier decomposition, and the sum of squares of the seasonal factors is equal to the sum of squares of the amplitudes of the harmonics.

#### 5.5.2.2.2 Model Results

The results of the final model used in testing for hidden seasonal patterns in the white maize data indicate that the random component should be left out of the model. This result appears to be common when fitting the different components to time series data (Harvey, Trimbur & Van Dijk, 2006:2). Even though a higher order cycle was used, the problem of implausible estimates remains.<sup>64</sup> The results of the free parameters are set out in Table 5.12.

<sup>64</sup> Making use of a higher order cycle produces a smoother extracted cycle (Harvey and Trimbur, 2003).

TABLE 5.12 FINAL ESTIMATES OF THE FREE PARAMETERS FOR WHITE MAIZE

Component	Parameter	Estimate	Std Error	t - Value	Pr >  t
Level	Error Variance	0.00038242	0.0012414	0.31	0.7582
Season	Error Variance	3.27E-12	2.40E-09	0.00	0.9987
Cycle	Damping Factor	0.96033	0.01877	51.16	<.0001
Cycle	Period	43.85977	10.58712	4.14	<.0001
Cycle	Error Variance	0.00761	0.0039904	1.91	0.0566

The final results regarding the model indicate that the level and seasonal components are not statistically significant at the 10 percent level of statistical significance. The cycle length was found to be 43.86 periods (months) with a damping factor of 0.9603. Both these parameters were found to be statistically significant at the 1 percent level of statistical significance.

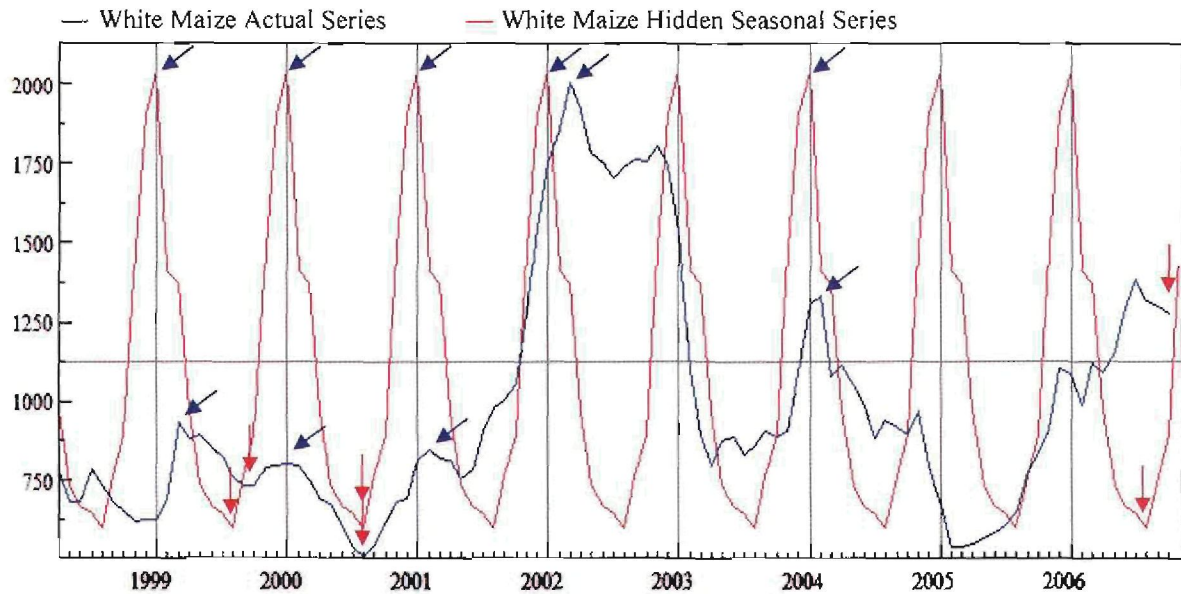
The fact that this model is sensitive to the influence of the noise between seasons can be responsible for this result. These problems could be overcome by testing for seasonal patterns over longer periods (Venter, 2007).

#### 5.5.2.2.3 Final Results – White Maize

Despite the fact that the results were not statistically significant they correlate well with the actual series. These results might be useful for the pairs trader as they give an indication of the underlying seasonal patterns in the data. Graph 5.7 illustrates the relationship between the actual monthly average spot price series and the hidden seasonal pattern for the white maize data.<sup>65</sup>

<sup>65</sup> Please note that the price of the maize series are given on the y-axis while the date is shown on the x-axis.

GRAPH 5.7 MONTHLY AVERAGE SPOT PRICE SERIES OF WHITE MAIZE VS HIDDEN SEASONAL



The hidden seasonal pattern extracted for the white maize series appears to be useful in signalling most of the directional changes in the actual series. The blue arrows in Graph 5.7 above shows that the UCM series leads the actual series into peaks, and the red arrows indicate the instances where the UCM series leads the actual series into troughs. Although this is not always the case, the hidden seasonal pattern appears to be a more accurate representation of the behaviour of the white maize series over the test period. The hidden seasonal pattern is also more consistent than the seasonal factor series discussed under the ARIMA-type modelling above. Because the UCM series is more consistent over time, it can be used with more confidence than the deterministic-type models. A further examination of these patterns will be presented in chapter 7. Section 5.5.2.3 that follows discusses the testing for hidden seasonal patterns in the yellow maize data.

### 5.5.2.3 Yellow Maize

#### 5.5.2.3.1 The Model

The final model used in testing for unobserved seasonal patterns in the yellow maize data took the form of:

$$y_t = \mu_t + \varphi_t + \gamma_t + \varepsilon_t, \quad (5.36)$$

where  $\mu_t$  is the trend,  $\varphi_t$  the cyclical component,  $\gamma_t$  the seasonal component, and  $\varepsilon_t$  the random component. The slope and the random components were once again not statistically significant at the 10 percent level of statistical significance with  $p$ -values of 0.430 and 0.999 respectively. They were therefore left out of the final model. For the final model, the trend component thus took the form of a random walk with noise (identical to the model used for white maize). The model for  $\mu_t$  is thus given by:

$$\mu_t = \mu_{t-1} + \eta_t \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2), \quad (5.37)$$

where the disturbance term  $\eta_t$  is assumed to be independent. The trend component  $\mu_t$  is a random walk with noise, and is modelled as a time-varying level component without any persistent upward or downward drift. The smoothed cyclical component option in SAS<sup>®</sup> (SAS, 2007a) was again used to model for hidden seasonality in the yellow maize data. The smoothed cycle specification used in the final model thus took the form of  $\psi_t = \psi_t^{(k)}$  where:

$$\begin{bmatrix} \psi_t^{(j)} \\ \psi_t^{*(j)} \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1}^{(j)} \\ \psi_{t-1}^{*(j)} \end{bmatrix} + \begin{bmatrix} \psi_t^{(j-1)} \\ \psi_t^{*(j-1)} \end{bmatrix}, \quad (5.38)$$

for  $j = 1, \dots, k$ ;  $t = 1, \dots, n$ ; and where  $v_t = \psi_t^{(0)}$  and  $v_t^* = \psi_t^{*(0)}$  are mutually uncorrelated white noise disturbances with zero means and common variance  $\sigma_v^2$ . Since  $\rho < 1$ ,  $\psi_t$  has a stationary distribution with zero mean and variance  $\sigma_v^2 / (1 - \rho^2)$ . Because the amount of observations used for yellow maize was the same as that used for white maize, the seasonal component was fixed and not allowed to be stochastic. The seasonal component thus took the trigonometric form, as was the case with the tests performed for white maize. As before  $\gamma_t$  is modelled as a sum of cycles of different frequencies and is given as (Harvey & Scott, 1994:1328):

$$\sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t} \quad (5.39)$$

where  $s$  equals  $(s-1)/2$ . The cycles  $\gamma_{j,t}$  have frequencies  $\lambda_j = 2\pi j/s$  and are specified by the matrix equation (SAS, 2006a:1633):

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \psi_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad (5.40)$$

where the disturbances  $\omega_{j,t}$  and  $\omega_{j,t}^*$  are assumed to be independent. Where  $j$  is fixed,  $\omega_{j,t}$  and  $\omega_{j,t}^* \sim N(0, \sigma_\omega^2)$ . The decomposition of the seasonal effects into these harmonics ( $\gamma_{j,t}$ ) is identical to its Fourier decomposition, and the sum of squares of the seasonal factors is equal to the sum of squares of the amplitudes of the harmonics.

### 5.5.2.3.2 Model Results

The results of the free parameters are set out in Table 5.13 below. As was the case with the white maize model, the slope and random components were left out of the model.

TABLE 5.13 FINAL ESTIMATES OF THE FREE PARAMETERS FOR YELLOW MAIZE

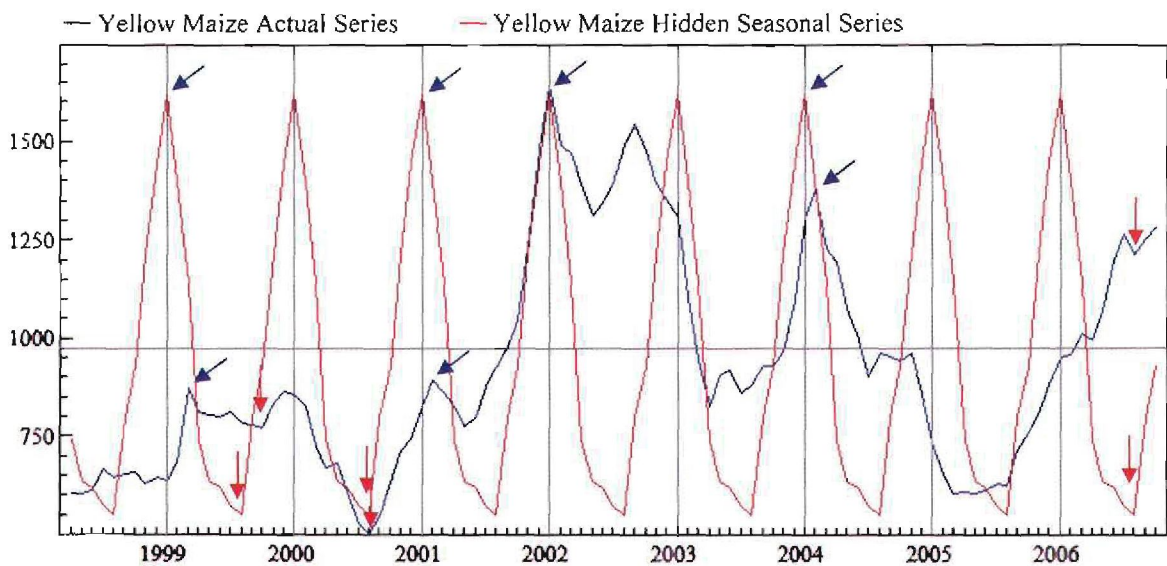
Component	Parameter	Estimate	Std Error	t - Value	Pr >  t
Level	Error Variance	0.00494	0.0009814	5.03	<.0001
Season	Error Variance	3.79E-13	5.59E-10	0.00	0.9995
Cycle	Damping Factor	0.99876	0.008734	114.35	<.0001
Cycle	Period	27.78888	1.08681	25.57	<.0001
Cycle	Error Variance	0.00005024	0.00005537	0.91	0.3642

The final results regarding the model indicate that the level component is statistically significant at the 1 percent level of statistical significance. The seasonal component is not statistically significant. The cycle length was found to be 27.79 periods (months) with a damping factor of 0.999. Both these parameters were found to be statistically significant at the 1 percent level of statistical significance.

### 5.5.2.3.3 Final Results – Yellow Maize

Even though the seasonal component was not found to be statistically significant, the results still correlate well with the actual series. Graph 5.8 illustrates the relationship between the actual monthly average spot price series and the hidden seasonal pattern for the yellow maize data.

GRAPH 5.8 MONTHLY AVERAGE SPOT PRICE SERIES OF YELLOW MAIZE VS HIDDEN SEASONAL



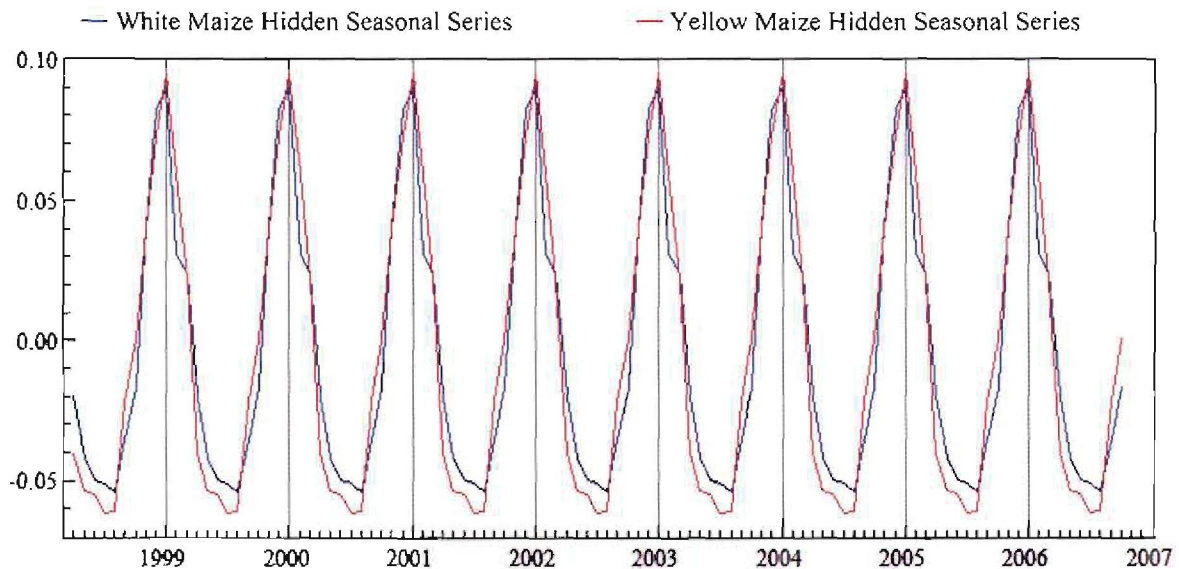
The unobserved seasonal pattern extracted for the yellow maize data appears to be useful in signalling the bulk of the changes in direction for the actual yellow maize series. Even though this is not always the case, the hidden seasonal pattern appears to be an accurate representation of the behaviour of the yellow maize series over the test period (these patterns will be discussed again in chapter 6).

### 5.5.2.4 White Maize vs. Yellow Maize

As previously mentioned, UCM can be utilised in various ways. Firstly, this stochastic method for testing seasonality may be used by the commodity pairs trader as a guide to the seasonal behaviour of the underlying commodities that he is trading, and secondly as an additional tool for selecting pairs. This can be deduced by considering the UCM results regarding the tests for hidden seasonality in both the white and yellow maize spot data. Graph 5.9 below compares the hidden seasonal patterns of the white and

yellow maize spot series as an example of how UCM was used in this research project, as described in chapter 6.<sup>66</sup>

GRAPH 5.9 WHITE MAIZE HIDDEN SEASONAL VS. YELLOW MAIZE HIDDEN SEASONAL



#### 5.5.2.5 Testing for Seasonality in the South African Maize Data with UCM

The tests discussed in sections 5.5.2.2 and 5.2.2.3 above were conducted for all of the maize series. The model used in testing for hidden seasonality in all the maize data took the form of:

$$y_t = \mu_t + \varphi_t + \gamma_t + \varepsilon_t. \quad (5.41)$$

where  $\mu_t$  is the trend,  $\varphi_t$  the cyclical component,  $\gamma_t$  the seasonal component, and  $\varepsilon_t$  the random component. As with the tests performed for the white and yellow maize series, preliminary tests were performed before specifying the final model. During the preliminary testing phase both the irregular and slope components were included in each of the models used for testing the individual maize series.<sup>67</sup> The results of the preliminary tests are reported in Table 5.14, indicating the significance of the various components.

<sup>66</sup> Graph 5.9 only serves as an example of the comparisons that will be reported in chapter 6 for all the possible pairs combinations. The hidden seasonal data is given on the y-axis, and the date on the x-axis.

<sup>67</sup> See sections 5.5.2.2 and 5.5.2.3 for an explanation of the procedure of selecting the correct UCM model.

TABLE 5.14 SIGNIFICANCE ANALYSES OF THE FREE PARAMETERS FOR THE MAIZE DATA (PRELIMINARY TESTS)

Agricultural Commodity Series	Component	DF	Chi-Square	Pr > ChiSq
WMSPOT	Irregular	1	0	1
	Level	1	4074.8	<.0001
	Slope	1	0.19	0.665
	Cycle	2	4.01	0.135
	Season	11	10.7	0.469
WM 1 away	Irregular	1	0	1
	Level	1	3915.25	<.0001
	Slope	1	0.17	0.683
	Cycle	2	3.44	0.179
	Season	11	9.57	0.57
WM 2 away	Irregular	1	0	1
	Level	1	3863.62	<.0001
	Slope	1	0.14	0.706
	Cycle	2	3.06	0.217
	Season	11	8.36	0.681
WM 3 away	Irregular	1	0	1
	Level	1	2460.64	<.0001
	Slope	1	2.52	0.113
	Cycle	2	0.23	0.891
	Season	11	9.77	0.551
YMSPOT	Irregular	1	0	1
	Level	1	10941.4	<.0001
	Slope	1	0.52	0.47
	Cycle	2	10.79	0.005
	Season	11	19.42	0.054
YM 1 away	Irregular	1	0	1
	Level	1	6840.54	<.0001
	Slope	1	0.54	0.462
	Cycle	2	6.03	0.049
	Season	11	13.96	0.235
YM 2 away	Irregular	1	0	1
	Level	1	5829.43	<.0001
	Slope	1	0.52	0.47
	Cycle	2	3.83	0.147
	Season	11	8.82	0.639
YM 3 away	Irregular	1	0	1
	Level	1	6099.1	<.0001
	Slope	1	0.62	0.429
	Cycle	2	2.31	0.315
	Season	11	11.07	0.437

From Table 5.14, it is clear that the  $p$ -values for the slope and irregular components are not statistically significant, and were therefore left out of the final models. The trend component for all of the final models for the individual maize series thus took the form of a random walk with noise. Hence, the model for  $\mu_t$  can be given by:

$$\mu_t = \mu_{t-1} + \eta_t \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2), \quad (5.42)$$

where the disturbance term  $\eta_t$  is assumed to be independent. The trend component  $\mu_t$  is a random walk (the level) and is thus modelled as a time-varying level component without any persistent upward or downward drift (no slope component is included).

As was the case for the white and yellow maize spot series, Harvey and Trimbur (2003) were followed in making use of the smoothed cycle option in SAS<sup>®</sup> (SAS, 2007a). The final model used in testing for unobserved seasonality in the maize data thus took the form of  $\psi_t = \psi_t^{(k)}$  where:

$$\begin{bmatrix} \psi_t^{(j)} \\ \psi_t^{*(j)} \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1}^{(j)} \\ \psi_{t-1}^{*(j)} \end{bmatrix} + \begin{bmatrix} \psi_t^{(j-1)} \\ \psi_t^{*(j-1)} \end{bmatrix}, \quad (5.43)$$

for  $j = 1, \dots, k$ ;  $t = 1, \dots, n$ ; and where  $v_t = \psi_t^{(0)}$  and  $v_t^* = \psi_t^{*(0)}$  are mutually uncorrelated white noise disturbances with zero means and common variance  $\sigma_v^2$ . In this case  $\rho < 1$ , and  $\psi_t$  thus has a stationary distribution with zero mean and variance  $\sigma_v^2 / (1 - \rho^2)$ .

The seasonal component takes the trigonometric form since there is not enough data to allow a changing pattern to be estimated. In this form  $\gamma_t$  is modelled as a sum of cycles of different frequencies and the model is given as (Harvey & Scott, 1994:1328):

$$\sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t}, \quad (5.44)$$

where  $s$  equals  $(s-1)/2$ . The cycles  $\gamma_{j,t}$  have frequencies  $\lambda_j = 2\pi j/s$  and are specified by the matrix equation (SAS, 2006a:1633):

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \psi_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad (5.45)$$

where the disturbances  $\omega_{j,t}$  and  $\omega_{j,t}^*$  are assumed to be independent. Where  $j$  is fixed, and where  $\omega_{j,t}$  and  $\omega_{j,t}^* \sim N(0, \sigma_w^2)$ .

The significance results of the final model used in testing for hidden seasonal patterns in the maize data are set out in Table 5.15, while the final estimates for each component are reported in Table 5.16.

TABLE 5.15 SIGNIFICANCE ANALYSES OF THE FREE PARAMETERS FOR THE MAIZE DATA (FINAL TESTS)

Agricultural Commodity Series	Component	DF	Chi-Square	Pr > ChiSq
WMSPOT	Level	1	2744.84	<.0001
	Cycle	2	5.15	0.076
	Season	11	10.08	0.524
WM 1 away	Level	1	2820.68	<.0001
	Cycle	2	4.01	0.134
	Season	11	8.98	0.624
WM 2 away	Level	1	3770.49	<.0001
	Cycle	2	3.09	0.214
	Season	11	8.47	0.671
WM 3 away	Level	1	2681.8	<.0001
	Cycle	2	1.51	0.469
	Season	11	9.67	0.56
YMSPOT	Level	1	10547.9	<.0001
	Cycle	2	11.06	0.004
	Season	11	19.63	0.051
YM 1 away	Level	1	6632.01	<.0001
	Cycle	2	6.47	0.039
	Season	11	14.21	0.221
YM 2 away	Level	1	5690.69	<.0001
	Cycle	2	4.26	0.119
	Season	11	8.97	0.625
YM 3 away	Level	1	5885.41	<.0001
	Cycle	2	2.7	0.259
	Season	11	11.15	0.431

TABLE 5.16 FINAL ESTIMATES OF THE FREE PARAMETERS FOR THE MAIZE DATA (FINAL TESTS)

Agricultural Commodity Series	Component	Parameter	Estimate	Approximate Std Error	t - Value	Approximate Pr >  t
WMSPOT	Level	Error Variance	0.001	0.002	0.430	0.669
	Season	Error Variance	0.000	0.000	0.000	0.999
	Cycle	Damping Factor	0.960	0.019	50.130	<.0001
	Cycle	Period	42.003	12.447	3.370	0.001
	Cycle	Error Variance	0.007	0.004	1.690	0.091
WM 1 away	Level	Error Variance	0.001	0.002	0.460	0.642
	Season	Error Variance	0.000	0.000	0.000	0.999
	Cycle	Damping Factor	0.961	0.019	50.620	<.0001
	Cycle	Period	44.238	12.404	3.570	0.000
	Cycle	Error Variance	0.007	0.004	1.730	0.084
WM 2 away	Level	Error Variance	0.006	0.002	2.730	0.006
	Season	Error Variance	0.000	0.000	0.000	0.998
	Cycle	Damping Factor	0.974	0.028	34.440	<.0001
	Cycle	Period	25.423	3.061	8.310	<.0001
	Cycle	Error Variance	0.001	0.001	1.330	0.182
WM 3 away	Level	Error Variance	0.001	0.002	0.460	0.648
	Season	Error Variance	0.000	0.000	0.000	0.998
	Cycle	Damping Factor	0.955	0.022	42.940	<.0001
	Cycle	Period	46.291	18.614	2.490	0.013
	Cycle	Error Variance	0.006	0.004	1.480	0.140
YMSPOT	Level	Error Variance	0.005	0.001	4.090	<.0001
	Season	Error Variance	0.000	0.000	0.000	0.999
	Cycle	Damping Factor	0.998	0.012	82.510	<.0001
	Cycle	Period	28.118	1.209	23.260	<.0001
	Cycle	Error Variance	0.000	0.000	0.920	0.358
YM 1 away	Level	Error Variance	0.004	0.001	2.760	0.006
	Season	Error Variance	0.000	0.000	0.000	0.999
	Cycle	Damping Factor	0.990	0.021	48.230	<.0001
	Cycle	Period	27.863	1.951	14.280	<.0001
	Cycle	Error Variance	0.000	0.000	1.120	0.264
YM 2 away	Level	Error Variance	0.003	0.001	2.400	0.017
	Season	Error Variance	0.000	0.000	0.000	0.999
	Cycle	Damping Factor	0.979	0.026	38.250	<.0001
	Cycle	Period	27.360	3.264	8.380	<.0001
	Cycle	Error Variance	0.001	0.001	1.280	0.201
YM 3 away	Level	Error Variance	0.003	0.002	2.210	0.027
	Season	Error Variance	0.000	0.000	0.000	0.999
	Cycle	Damping Factor	0.971	0.031	31.260	<.0001
	Cycle	Period	27.093	4.176	6.490	<.0001
	Cycle	Error Variance	0.001	0.001	1.270	0.203

The final results indicate that the longest cycle length is recorded for the WM 3 away series (46.291 months), while the shortest cycle is recorded for the WM 2 away series (25.423 months). All the maize series' cycle parameters were found to be statistically significant at the 1 percent level of statistical significance.

The seasonal components are not statistically significant at the 10 percent level of statistical significance for any of the maize series. This is a peculiar result, since all of the seasonal patterns produced during the UCM process appear to move with the original series for the different maize series. However, this outcome may be the result of the sensitivity of the model to the influence of the noise. According to Venter (2007) these problems might be overcome by testing for seasonal patterns over longer periods.

### 5.5.3 Conclusion

When considering the hidden seasonal patterns in the agricultural commodities data in graphs 5.7 and 5.8 above, it is evident that these patterns correlate with the actual spot series. The UCM method appears able to explain the seasonal patterns in the data well enough for the trader to take these patterns into consideration when trading.

The seasonal patterns found for both white and yellow maize correlate with the fundamental factors that play a role in the agricultural commodities market. The peak in the hidden seasonal pattern over January for both white and yellow maize is explained by the increased uncertainty about the weather (drought) over the growing season. The trough shows the drop in prices over July/August, which is the middle of the delivery period. The reason for this is that there is greater certainty of the size of the yield during this period.

Since UCM explains the actual data so well it can be utilised as an additional tool to be used by the commodities pairs trader. It can serve as a measure of the seasonality in the maize data, and as a tool to validate pairs during the pairs selection process. To this end, the strong statistical relationship between the various maize series' seasonal patterns is an indication that their prices will mean revert after moving in opposite directions. This quality will ensure that the selected pairs can be traded successfully.

## 5.6 Summary

The testing of these seasonal patterns, although not the main focus of this thesis, is important as the results are useful for the pairs trading strategy that this project implements. Seasonality tests are important for two reasons. Firstly, the information regarding these seasonal patterns in the data is indispensable when compiling a pairs portfolio. By comparing the seasonal patterns in the various commodity price series, the pairs trader will either accept or reject a pair on the basis of the correlation between their seasonal patterns. Secondly, knowledge about the seasonal patterns in the data also allows the trader to make informed decisions about the future behaviour of these commodity prices.<sup>68</sup>

In testing for seasonal patterns in agricultural commodities time series data two approaches were used: a deterministic and a stochastic approach. Tests for the existence of seasonal patterns were first performed by making use of GARCH models. However, this deterministic model did not render satisfactory results. Even though the results were largely statistically significant, no meaningful relationship between the results and the actual time series was found.

The second deterministic model was performed in the form of the ARIMA-based X12 procedure. These tests indicated that even though stable seasonality was found no identifiable seasonality could be found in the data. When comparing the Henderson-curve of the white maize spot series to the original white maize series, the Henderson-curve does well in following the original series for the entire study period. The Henderson-curve of the yellow maize spot series does equally well in following the original yellow maize spot series.

Comparison of the seasonal factors series of the white maize series to the actual white maize series indicates that the seasonal factors series leads the original series for the majority of the data period. The same results were reported for the seasonal factors series of the yellow maize data.

---

<sup>68</sup> Although the hidden seasonal patterns were not used explicitly in tests, these patterns should be kept in mind when assembling an agricultural commodity pairs portfolio.

Tests were also performed using a stochastic approach in the form of the stochastic UCM method. This model tests for hidden or unobserved seasonality in time series data. Although no statistically significant results were reported, these patterns seem to describe the actual series more accurately than the two deterministic models (GARCH and X12). In several cases the unobserved seasonal patterns seem to be a leading indicator to the movements of the actual series.

These seasonal patterns in the agricultural commodities data are important for the trader considering an agricultural commodity pairs portfolio. If the trader has an idea of the movement of the commodities prices in his pairs portfolio, it makes the decisions of when to change positions easier. Knowledge about the seasonal patterns in the data is not solely crucial in terms of the general trend of the commodity prices, but will also be useful in identifying pairs during the pairs selection process.

# CHAPTER 6

## The Pairs Selection Process

*We may know many things, and the more we know, the better we are placed to make the right decisions, but knowledge alone is not a sufficient basis for making decisions*

George Soros  
(Soros, 2003:3)

### 6.1 Introduction

Traditionally, pairs have been identified by testing for the statistical relationship between two commodities (or shares). These statistical measures typically include tests for correlation (both Pearson's correlation, Spearman's rank correlation) and Beta (Ehrman, 2006). In order to identify pairs in the maize data, these traditional measures were used to test for the statistical relationships between various combinations of white and yellow maize series. In addition to this, co-integration was included as a statistical measure in the array of tests to be performed during the pairs selection process.

To successfully select a commodity pairs portfolio, the pairs trader will have to pair commodities that have a strong statistical relationship in the long-run. Tests for co-integration, therefore, can be vital to verifying such a relationship. Pearson's correlation and Spearman's rank correlation were utilised to validate the co-integration results regarding the long-run relationship between two commodities. Beta was employed as an indicator of the co-movement of the two maize series comprising the pair (this use of Beta will be further discussed in chapter 7).

Pairs were also assembled using the seasonality tests as an additional pairs selection tool, as has been explained in chapter 5. Two seasonality tests were performed to this end. The first test compared the Henderson-curves of the two series in terms of the correlation coefficient between the Henderson-curves, in order to test for the co-movement between their long-run trends. The second test compared the two commodities' seasonal factors series in terms of the correlation coefficient between the seasonal factors series, in order to get an idea of the short-term co-movement of the data

series.<sup>69</sup> This approach to pairs selection is based on the theory that two entities can be paired, if they reveal similar seasonal patterns. A pair, therefore, would be selected if the two entities' seasonal patterns were co-moving over time.

For the purpose of this project, two commodity series were only considered a pair once they had been deemed a pair according to all the pairs selection criteria discussed above as well as according to the traditional pairs selection tools discussed below. Once specific pairs were identified, they were traded as unitary entities by making use of fundamental analysis and several technical trading analysis tools, as has been discussed in chapter 2 and chapter 3 respectively.

Since the data used in this chapter are the same as the data used in the chapter 5, no additional explanation of the data is provided here.<sup>70</sup> Furthermore, it is to be noted that each of the series as has been described in chapter 5 were converted to a Henderson-curve (trend-cycle), a seasonal factors series, and a hidden seasonal series. These series were used to compare the seasonal relationships between the various commodity series. The statistical relationship between the price series that has been described in chapter 5 were tested by making use of several statistical procedures, which are discussed below.

In section 6.2, the methodology followed in the pairs selection process by means of statistical measures is explained. Next, section 6.3 reports on the use of deterministic seasonality measures to verify the validity of the pairs already selected by means of the statistical tests. Finally, section 6.4 reports on the use of stochastic seasonality measures to verify the validity of the pairs already selected by both the statistical and the deterministic seasonality tests.

## 6.2 Pairs Selection: Statistical Measures

### 6.2.1 Introduction

It is to be noted that two commodities were only considered a pair if they met all the statistical criteria. The first step was to establish which pairs have a co-integrating

---

<sup>69</sup> It is to be noted that the pairs trader will also be able to use the results of these tests as a guide regarding the seasonal patterns in the data.

<sup>70</sup> Please refer to section 5.2.2 for a discussion on the data used.

relationship between them. Thereafter, steps 2 and 3 were tests performed to establish the strength of the two commodities' statistical relationship. This was done by testing for Pearson's correlation coefficient (step 2), followed by tests for Spearman's rank correlation coefficient (step 3). During step 4 those pairs that still met all of the criteria of each step were subjected to tests, which establish the Beta value between the two commodities that make up each pair. After performing these steps, the elimination process continued by testing the co-movement of the commodities' seasonal patterns (steps 5 and 6). This was done by finding the correlating relationships between the different series' Henderson-curves (step 5) and the correlating relationships between the seasonal factors series (step 6). The elimination process was concluded by testing the co-movement of the commodities' hidden seasonal patterns during the seventh and final step.

Even though none of the methods discussed below are new, no record exists of co-integration, Henderson-curves, seasonal factors series, and hidden seasonal patterns being used to identify pairs.

This section first explains the various statistical measures used in testing whether two agricultural commodities can be traded as a pair. After explaining what these statistical measures are and how they work, their role in the commodity selection process is examined.

### 6.2.2 Statistical Measures

To test for commodity pairs, it is necessary to study the statistical measures or indicators that are to be used during the selection of commodity pairs. Since agricultural commodities pairs were selected based on the long-run statistical relationship between their prices, Grangers' co-integration is discussed first. This section also discusses Pearson's correlation and Spearman's rank correlation as methods of establishing a price relationship between two commodities. Beta is discussed here as well, though it was not used to determine whether a particular pair was to be considered in the final portfolio. Rather it was utilised during the test-trading process as an indication of the co-movement between the two maize series comprising a pair, as will be reported in chapter 7.

### 6.2.2.1 Co-integration

The concept of co-integration was first developed by Granger (1981) and later improved on by Engle and Granger (1987). This statistical measurement tool integrates short-run dynamics with long-run equilibria without the downside of losing potentially important information regarding long-run relationships. The testing of co-integration is discussed in more detail in section 6.2.2.2.4 below.

In order to explain co-integration, the concepts of stationarity and unit roots are first explained. Consider the simple autoregressive equation by way of example:

$$R_t = \alpha + \phi R_{t-1} + e_t. \quad (6.1)$$

When subtracting  $R_{t-1}$  on both sides of equation 6.3 it yields:

$$\Delta R_t = \alpha + \rho R_{t-1} + e_t, \quad (6.2)$$

where  $\rho = \phi - 1$ . If  $\phi = 1$ , then  $\rho = 0$ , which will cause  $\Delta R_t$  to fluctuate randomly around  $\alpha$ . In the case where  $|\phi| < 1$ ,  $R$  will be stationary and where  $\phi = 1$ ,  $R$  will be non-stationary (unit root). Since  $R$  has a unit root its autocorrelations will be close to one and will not reduce much as the lag length increases.  $R$  will possess a 'long memory' in that any impact caused by a shock on  $R_t$  will be present for a long time after the event.

In the case of  $R$  having a unit root, the series  $R_t$  will exhibit trend behaviour and  $\Delta R$  will be stationary (Koop, 2005:145). The time series  $R_t$  is further said to be integrated of an order 1, thus  $I(1)$  when  $\Delta R_t$  is a stationary time series; and it is  $I(0)$  when the series  $R_t$  itself is stationary.<sup>71</sup> Where  $\Delta R_t$  is  $I(1)$  the series  $R_t$  will be integrated  $I(2)$  (Maddala, 2002:556). Now consider the equation:

$$y_t = \alpha + \beta x_t + \gamma z_t + \varepsilon_t, \quad (6.3)$$

---

<sup>71</sup> Differencing a variable with a unit root will remove the unit root.

where  $y_t$  is the dependent variable,  $x_t$  and  $z_t$  are the exogenous regressors, and  $\varepsilon_t$  will be a white-noise mean-zero sequence. Granger (1981:121) argues that an equation has to be consistent in order to be meaningful, in the sense that “a simulation of the explanatory right-hand side should produce the major properties of the variable being explained”. For  $y_t$  to be a non-seasonal variable, and  $\varepsilon_t$  to be white noise, the term  $\beta x_t + \gamma z_t$  has to be non-seasonal if  $x_t$  and  $z_t$  in equation 6.3 above, are seasonal (variables with unit-roots). This would only be possible with the incorporation of a constraint  $f = c/b$  such that the seasonal component in  $x_t$  is exactly the reverse of  $f$  times the seasonal component in  $z_t$ .

If variable  $z_t$  can be made approximately stationary by differencing it  $d$  times, it is called integrated of order  $d$ , or  $I(d)$ . Weakly stationary random variables are thus  $I(0)$ . Further, it is true that if  $z_t \sim I(1)$ , then  $\Delta z_t \sim I(0)$  (KVA, 2003:5).

It is to be noted that  $I(1)$  variables dominate  $I(0)$  variables. In a linear combination of variables the variation of  $I(1)$  variables overwhelms the variation of  $I(0)$  variables (KVA, 2003:5). The reason for this occurrence is that the theoretical infinite variance from an  $I(1)$  series is entirely derived from the contribution of the low frequencies or long-run part of the series. Compared to an  $I(0)$  series, an  $I(1)$  series has dominant long swings, thus causing the series to be rather smooth. Because of the relative sizes of the variances, it is always true that the sum of an  $I(1)$  series and an  $I(0)$  series will be  $I(1)$ . Also, if  $\alpha$  and  $\beta$  are constants,  $\beta \neq 0$ , and  $x_t \sim I(d)$ , then  $\alpha + \beta x_t$  will also be  $I(d)$ . If, for the equation:

$$z_t = x_t - \alpha y_t, \tag{6.4}$$

$x_t$  and  $y_t$  are both  $I(1)$ , then it is usually true that equation 6.4 will also be  $I(1)$  (Engle & Granger, 1987:253). Following from the discussion above, it is to be expected that the resulting error will also exhibit non-stationary behaviour in the case where two variables containing unit roots (seasonal patterns) are regressed on one another. If the series are co-integrated, however, the resulting error will be stationary. For the time series regression (Maddala, 2002:556):

$$y_t = \alpha x_t + z_t, \quad (6.5)$$

$x_t$  and  $y_t$  are thus co-integrated where  $x_t \sim I(1)$ ,  $y_t \sim I(1)$ , and  $y_t - \alpha x_t$  is  $I(0)$ . When two variables are co-integrated it means that there exists a long-run relationship between them of such a nature that they will not drift too far apart from each other over time (Engle & Granger, 1987:253).

Since co-integration effectively confirms the long-run relationship between two variables over time, it is instrumental in identifying pairs. Once the trader is confident that two price series will move together over time, he can safely assume that the gap between them will open and close regularly. This assurance will allow the trader to take a position with the certainty that the relationship between his long and short positions will always converge at a future point in time.

To test for co-integration, the series was first tested to ascertain whether there are any unit roots present in the data. Once unit roots were found, tests for co-integration were performed on those series that are difference stationary for the same order of difference. The following section discusses the methodology and results of the tests performed for finding long-run pairs.

### 6.2.2.2 Utilising Co-integration in Selecting Long-run Pairs

#### 6.2.2.2.1 Introduction

A necessary condition for co-integration is that each of the variables must be integrated of the same order (Granger, 1986). In order to investigate the stationary property of each variable, the Augmented Dicky-Fuller (ADF) and Phillips Perron (PP) unit roots technique was used. The ADF test allows for more dynamics in the DF regression and, therefore, is over-parameterised in the case of the first-order case, but correctly specified for higher order cases. Phillips and Perron (1988) also provide a test (PP test) for co-integration with the intended benefit of taking care of moving average errors. It is well known that the co-integration tests are very sensitive to the choice of lag length.

The Schwartz Information Criterion (SIC) is used to select the number of lags required in the co-integration test. Sections 6.2.2.2.2 and 6.2.2.2.3 discuss the methodology followed and results obtained in testing for unit roots, while sections 6.2.2.2.4 and 6.2.2.2.5 discuss the methodology followed and results obtained in testing for co-integration.

#### 6.2.2.2.2 Testing for Unit Roots: Methodology

Before testing for a co-integrating relationship between these series, all the series were first tested to ascertain whether there are any unit roots present in the data. In testing for the existence of unit roots in the data, the ADF and PP methods in EViews<sup>®</sup> were used (QMS, 2004b). The null hypothesis of both the ADF and PP tests states that the data contain a unit root. The null hypothesis for tests performed for unit roots is thus given by:  $H_0 : \rho = 1$ , against the alternative of  $H_1 : \rho < 1$ .

The ADF and PP tests for unit roots were performed simultaneously. The first set of tests were performed on the original data series (before differencing) without specifying any exogenous variables. If unit roots were found to be present in the data, differences were taken in order to determine the order of integration.<sup>72</sup> Differences were taken and the tests repeated until the data was found to be stationary.

All the series integrated to the same order were then tested for a co-integrating relationship between them. After testing for unit roots, tests for co-integration were performed on those series that are difference stationary for the same order of difference.

#### 6.2.2.2.3 Testing for Unit Roots: Results

The test results indicate that all the series are non-stationary (with  $p$ -values exceeding the 10 percent level of statistical significance). Before taking first differences the ADF test revealed that all the series tested do indeed contain unit roots. All tests for unit roots were performed in EViews<sup>®</sup>.

---

<sup>72</sup> See section 6.4.2.1 for the discussion on the order of integration.

Table 6.1 below reports the results of the unit root tests performed by making use of ADF for the ordinary data series, while Table 6.2 reports on the results of the PP tests for the same data. The second set of tests were performed after taking first differences for all the series, and allowed for the inclusion of a deterministic trend. Tables 6.3 and 6.4 report the results of the second set of unit root tests (making use of ADF and PP respectively) for the first differences data series.

TABLE 6.1 UNIT ROOT TEST OF THE ORIGINAL SERIES – ADF

Agricultural Commodity Series	Lag Length	ADF test statistic	Critical value 95 %	Critical value 99 %	Prob. ( <i>p</i> -value)
White Maize Spot	2	-1.411	-2.863	-3.433	0.5785
White Maize 1 Contract away	1	-1.483	-2.863	-3.433	0.5425
White Maize 2 Contracts away	1	-1.415	-2.863	-3.433	0.5762
White Maize 3 Contracts away	1	-1.463	-2.863	-3.433	0.5524
Yellow Maize Spot	2	-1.964	-2.863	-3.433	0.3032
Yellow Maize 1 Contract away	2	-1.545	-2.863	-3.433	0.5104
Yellow Maize 2 Contracts away	0	-1.541	-2.863	-3.433	0.5127
Yellow Maize 3 Contracts away	2	-1.463	-2.863	-3.433	0.5780

TABLE 6.2 UNIT ROOT TEST OF THE ORIGINAL SERIES – PP

Agricultural Commodity Series	Bandwidth	PP test statistic	Critical value 95 %	Critical value 99 %	Prob. ( <i>p</i> -value)
White Maize Spot	14	-1.566	-2.863	-3.433	0.4996
White Maize 1 Contract away	1	-1.347	-2.863	-3.433	0.6097
White Maize 2 Contracts away	12	-1.475	-2.863	-3.433	0.5464
White Maize 3 Contracts away	12	-1.523	-2.863	-3.433	0.5219
Yellow Maize Spot	8	-1.926	-2.863	-3.433	0.3203
Yellow Maize 1 Contract away	1	-1.577	-2.863	-3.433	0.4940
Yellow Maize 2 Contracts away	0	-1.541	-2.863	-3.433	0.5127
Yellow Maize 3 Contracts away	13	-1.541	-2.863	-3.433	0.5128

TABLE 6.3 UNIT ROOT TEST OF THE FIRST DIFFERENCES – ADF

Agricultural Commodity Series	Lag Length	ADF test statistic	Critical value 95 %	Critical value 99 %	Prob. (p-value)
White Maize Spot	1	-31.395	-3.412	-3.963	0.0000
White Maize 1 Contract away	0	-37.647	-3.412	-3.963	0.0000
White Maize 2 Contracts away	0	-38.485	-3.412	-3.963	0.0000
White Maize 3 Contracts away	0	-36.691	-3.412	-3.963	0.0000
Yellow Maize Spot	1	-33.205	-3.412	-3.963	0.0000
Yellow Maize 1 Contract away	1	-32.804	-3.412	-3.963	0.0000
Yellow Maize 2 Contracts away	0	-42.726	-3.412	-3.963	0.0000
Yellow Maize 3 Contracts away	1	-32.156	-3.412	-3.963	0.0000

TABLE 6.4 UNIT ROOT TEST OF THE FIRST DIFFERENCES – PP

Agricultural Commodity Series	Bandwidth	PP test statistic	Critical value 95 %	Critical value 99 %	Prob. (p-value)
White Maize Spot	11	-38.819	-3.412	-3.963	0.0000
White Maize 1 Contract away	8	-37.453	-3.412	-3.963	0.0000
White Maize 2 Contracts away	9	-38.584	-3.412	-3.963	0.0000
White Maize 3 Contracts away	7	-36.803	-3.412	-3.963	0.0000
Yellow Maize Spot	6	-41.687	-3.412	-3.963	0.0000
Yellow Maize 1 Contract away	8	-39.917	-3.412	-3.963	0.0000
Yellow Maize 2 Contracts away	4	-42.669	-3.412	-3.963	0.0000
Yellow Maize 3 Contracts away	11	-40.390	-3.412	-3.963	0.0000

From the results it is evident that each price series is non-stationary in levels and stationary in first differences, indicating that all the data series are  $I(1)$ . Since all the series are  $I(1)$ , the next step was to test for co-integration between all the possible combinations of the data series.

If co-integrating relationships were found between the data series tested, it was accepted that there is a long-run relationship between the two series of such a nature that they will not drift too far apart over time. After this long-run relationship between the series was established, they were treated as long-run pairs and tested for the stability of their statistical relationship over time by making use Pearson's correlation, Spearman's rank correlation and Beta.

#### 6.2.2.2.4 Testing for Co-integration: Methodology

In order to utilise co-integration as a tool in the pair selection process, it is imperative to find an effective way to test for co-integration. Before testing for co-integration, tests must first be performed on the individual variables to assess whether they have unit roots. If the variables contain unit roots, the one must be regressed on the other so that unit root tests can be performed on the residuals. The two variables will be co-integrated if the residuals do not contain any unit roots (Maddala and Kim, 2000:40).

There are various ways in which the existence of a co-integrating relationship between variables can be tested. The earlier tests for co-integration, introduced by Engle and Granger (1987), are residual-based and include: the Durbin Watson (DW) test for stationary residuals, the DF test for the existence of a unit root, and the ADF test.

Several econometricians, however, are critical of the abovementioned methods of testing for co-integration between time series variables. Although Engle and Granger (1987) suggest using the DF and ADF tests to test for the existence of unit roots in the residuals of a regression of one variable on another, Maddala and Kim (2000:45) and Maddala (2002:549) advise against using these tests to test for the existence of unit roots. They state that the DF and ADF tests should not be considered for unit root testing because of their low power and size distortions (Maddala & Kim, 2000:78).

This critique of low power also extends to the PP test for the existence of unit roots in time series. As with the DF and ADF tests, Phillips and Perron's tests are also claimed to suffer from size distortions in finite samples when the data-generating process has a predominance of negative autocorrelations in first-differences (Maddala & Kim, 2000:81).

Despite this criticism, recent research suggests that the Johansen's test for co-integration remains the preferred test to use. When making use of Johansen's test for co-integration, ADF unit root tests were still used to establish whether the series are all  $I(1)$  before testing for co-integration. The PP test for the existence of unit roots in time series is also still widely used (see for example, Chang, Nieh & Wei, 2006; Islam & Hasan, 2006; and McMillan, 2005).

When making use of Johansen's maximum likelihood method, the null hypothesis is that there is no co-integration, against the alternative that co-integration is present. Johansen (1988) considers the simple case where  $X_t$  is integrated to the order 1, such that the first difference of  $X_t$  is stationary. If the process  $X_t$  is defined by an unrestricted VAR system of order  $(n \times 1)$  it is possible to express it as a vector error correction model. This model will take the form of:

$$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \dots + \Pi_k X_{t-k} + u_t, \quad (6.6)$$

where  $X_t = (n \times 1)$  is the vector of  $I(1)$  variables,  $\Pi_i = (n \times n)$   $i = 1, 2, \dots, k$  is the matrix of unknown parameters to be estimated,  $u_t$  is the independent and identically distributed (i.i.d)  $(n \times 1)$  vector of error terms, and  $t = 1, 2, \dots, m$  observations. By making use of  $\Delta = (1 - L)$  where  $L$  is the lags operator, the system described in equation 6.6 above can be parameterised in the error correction form as (Johansen & Juselius, 1990:170):

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + u_t, \quad (6.7)$$

where:

$$\Gamma_i = \sum_{j=1}^{k-1} \Pi_j - I \quad i = 1, 2, \dots, k-1, \quad (6.8)$$

and

$$\Pi = \sum_{j=1}^k \Pi_j - I. \quad (6.9)$$

and where  $\Delta X_t$  is an  $I(0)$  vector, and  $I$  is a  $(n \times n)$  identity matrix. Johansen's (1988) approach derives the maximum likelihood estimators of the co-integrating vectors for the autoregressive process with independent errors. The  $(n \times n)$  matrix  $\Pi$  can be written as the product of the two matrices  $\alpha$  and  $\beta$ , each of rank  $r$ , such that  $\Pi = \alpha\beta'$ ,

where  $\beta$  represents the  $r$  co-integrating vectors, and  $\alpha$  represents the matrix of the weighting elements (Johansen & Juselius, 1990:169). It is thus possible to rewrite equation 6.7 above as:

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + (\alpha\beta') X_{t-k} + u_t. \quad (6.10)$$

The maximum likelihood approach enables the testing of the hypothesis of  $r$  co-integrating relations between the elements of  $X_t$ :

$$H_0 : \Pi = \alpha\beta', \quad (6.11)$$

where the null hypothesis of no co-integrating relations,  $r = 0$ , implies that  $\Pi = 0$ . Testing for co-integration thus involves testing for the rank  $r$  of the  $\Pi$  matrix by examining whether the eigenvalues of  $\Pi$  are significantly different from zero. Johansen and Juselius (1990) propose two test statistics for evaluating the number of co-integrating vectors (or the rank of  $\Pi$ ): the trace ( $Tr$ ) statistic and the maximum eigenvalue (L-max) statistic. The likelihood ratio statistic for the trace test is (Johansen, 1991:1555):

$$-2 \ln Q = -T \sum_{i=r+1}^{p-2} \ln(1 - \hat{\lambda}_i), \quad (6.12)$$

where  $\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_p$  are the estimated  $p-r$  smallest eigenvalues. The null hypothesis is that there are at most  $r$  co-integrating vectors. In each case, the null hypothesis is tested against the general alternative. It is to be noted that the rejection of the null  $r \leq s-1$  leads to the conclusion that  $r \geq 1$ , when making use of the trace test. Seeing as the next null in sequence is  $r \leq s$ , and  $r \leq s-1$  has already been rejected, the new null is reduced to  $r = s$ . This process continues until the null is not rejected (Burke & Hunter, 2005:102).

The alternative to the likelihood ratio statistic for  $Tr$  given in equation 6.12, the L-max statistic is:

$$-2 \ln Q = -T \ln(1 - \hat{\lambda}_{r+1}). \quad (6.13)$$

In this test, the null hypothesis of  $r$  co-integrating vectors is tested against the alternative  $r+1$  co-integrating vectors. Thus, the null hypothesis  $r=0$  is tested against the alternative that  $r=1$ , and  $r=1$  against the alternative of  $r=1$ , and so forth (Burke & Hunter, 2005:100).

#### 6.2.2.2.5 Testing for Co-integration: Results

The tests for co-integrating relationships between various combinations of the data series were performed by making use of Johansen's test for co-integration in EViews® (QMS, 2004b). There are several options for the inclusion of a deterministic trend when testing for co-integrating relationships between two data series with Johansen's test for co-integration. The first option is to exclude a deterministic trend from the level data and not include any intercepts in the co-integrating equations:

$$H_2(r): \Pi y_{t-1} + Bx_t = \alpha \beta' y_{t-1}. \quad (6.14)$$

The second option is to exclude a deterministic trend from the level data but to include intercepts in the co-integrating equations:

$$H_1^*(r): \Pi y_{t-1} + Bx_t = \alpha (\beta' y_{t-1} + \rho_0). \quad (6.15)$$

The third option is to include a linear trend in the level data but to exclude intercepts from the co-integrating equations:

$$H_1(r): \Pi y_{t-1} + Bx_t = \alpha (\beta' y_{t-1} + \rho_0) + \alpha_{\perp} \gamma_0. \quad (6.16)$$

The fourth option is to include linear trends in both the level data and the co-integrating equations:

$$H^*(r): \Pi y_{t-1} + Bx_t = \alpha (\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_{\perp} \gamma_0. \quad (6.17)$$

The fifth option is to include quadratic trends in the level data and linear trends in the co-integrating equations:

$$H(r): \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_{\perp}(\gamma_0 + \gamma_1 t). \quad (6.18)$$

The terms associated with  $\alpha_{\perp}$  are the deterministic terms outside the co-integrating relations. According to Johansen (1995a),  $\alpha_{\perp}$  is the null space of  $\alpha$  such that  $\alpha' \alpha_{\perp} = 0$ . EViews<sup>®</sup> identifies the part inside the error correction term by regressing the co-integrating relations  $\beta' y_t$  on a constant, linear trend (QMS, 2004b).

The first option regarding the specification of the trends above should only be used in cases where the series has a zero mean. The second option should be used in cases where none of the series has a trend, while the fourth option is applicable in cases where the series are trend stationary. The fifth option does not produce plausible out-of-sample forecasts. Since the third option for trending series models the data the best, this option was used for testing for co-integrating relationships in the maize series.

The data used for the co-integration tests consists of the daily (end-of-the-day) logarithmic prices for a period stretching from the 26th of March 1998 until the 31st of October 2006 for both white and yellow maize (SAFEX, 2007). There are 2142 observations for both white and yellow maize. The results of the co-integration tests are given in Table 6.5 report on the tests performed. These tests allowed for trending data and inserted one lag for the first difference terms in the vector error correction model (VEC).

TABLE 6.5 JOHANSEN'S TEST FOR CO-INTEGRATION – LONG-RUN PAIRS

Agricultural Commodity Pair	Eigenvalue	Trace statistic	Critical value 99%	Prob. ( <i>p</i> -value)	Max-Eigen Statistic	Critical value 99%	Prob. ( <i>p</i> -value)
WMSPOT/WM 1 away	0.0135	29.650	19.937	0.000**	27.492	18.520	0.000**
WMSPOT/WM 2 away	0.0088	20.322	19.937	0.009**	17.981	18.520	0.012*
WMSPOT/WM 3 away	0.0128	29.257	19.937	0.000**	26.154	18.520	0.000**
WM 1 away/ WM 2 away	0.0151	32.820	19.937	0.000**	30.823	18.520	0.000**
WM 1 away/ WM 3 away	0.0157	34.290	19.937	0.000**	31.996	18.520	0.000**
WM 2 away/ WM 3 away	0.0158	34.657	19.937	0.000**	32.233	18.520	0.000**
YMSPOT/YM 1 away	0.0216	46.861	19.937	0.000**	43.876	18.520	0.000**
YMSPOT/YM 2 away	0.0132	29.673	19.937	0.000**	26.767	18.520	0.000**
YMSPOT/YM 3 away	0.0125	28.380	19.937	0.000**	25.221	18.520	0.001**
YM 1 away/ YM 2 away	0.0104	23.901	19.937	0.002**	21.239	18.520	0.003**
YM 1 away/ YM 3 away	0.0100	23.137	19.937	0.003**	20.276	18.520	0.005*
YM 2 away/ YM 3 away	0.0224	48.356	19.937	0.000**	45.770	18.520	0.000**
WMSPOT/YMSPOT	0.0098	23.704	19.937	0.002**	19.814	18.520	0.006**
WMSPOT/YM 1 away	0.0075	17.649	15.495	0.023*	15.163	14.265	0.036*
WMSPOT/YM 2 away	0.0067	15.626	15.495	0.048*	13.529	14.265	0.065
WMSPOT/YM 3 away	0.0080	18.214	15.495	0.019*	16.101	14.265	0.025*
WM 1 away/YMSPOT	0.0119	26.610	19.937	0.001**	23.977	18.520	0.001**
WM 1 away/ YM 1 away	0.0092	21.117	19.937	0.006**	18.749	18.520	0.009**
WM 1 away/ YM 2 away	0.0100	22.537	19.937	0.004**	20.417	18.520	0.005**
WM 1 away/ YM 3 away	0.0111	24.303	19.937	0.002**	22.500	18.520	0.002**
WM 2 away/YMSPOT	0.0095	21.576	19.937	0.005**	19.113	18.520	0.008**
WM 2 away/ YM 1 away	0.0079	18.199	15.495	0.019*	16.156	14.265	0.025*
WM 2 away/ YM 2 away	0.0095	21.350	19.937	0.006**	19.379	18.520	0.007**
WM 2 away/ YM 3 away	0.0113	24.737	19.937	0.002**	22.959	18.520	0.002**
WM 3 away/YMSPOT	0.0097	22.778	19.937	0.003**	19.512	18.520	0.007*
WM 3 away/ YM 1 away	0.0084	19.611	15.495	0.011*	17.042	14.265	0.018*
WM 3 away/ YM 2 away	0.0097	21.962	19.937	0.005**	19.656	18.520	0.006**
WM 3 away/ YM 3 away	0.0109	24.013	19.937	0.002**	22.102	18.520	0.002**

\* Denotes a co-integrating relationship at the 0.05 level, while \*\* denotes the existence of a co-integrating relationship at the 0.01 level

The Johansen test for co-integration indicates that all the combinations of maize series tested contain a co-integrating relationship. According to the results, the trader will be able to take a position with the certainty that the relationship between his long and short positions will always converge at a future point in time. Even though these pairs were found to have a long-run relationship over time, these results should only guide the trader in deciding which of the proposed pairs should undergo the next phase of testing.

#### 6.2.2.2.6 Conclusion

Since all the different combinations tested for co-integration have strong long-run statistical relationships, they can be considered long-run pairs. The next step was to test for the robustness of the statistical relationship of each of the pairs identified by the co-integration tests that were described in the preceding sections. This was achieved by first testing for the statistical relationships between the various pairs by means of Pearson's correlation coefficient, and Spearman's rank correlation.<sup>73</sup>

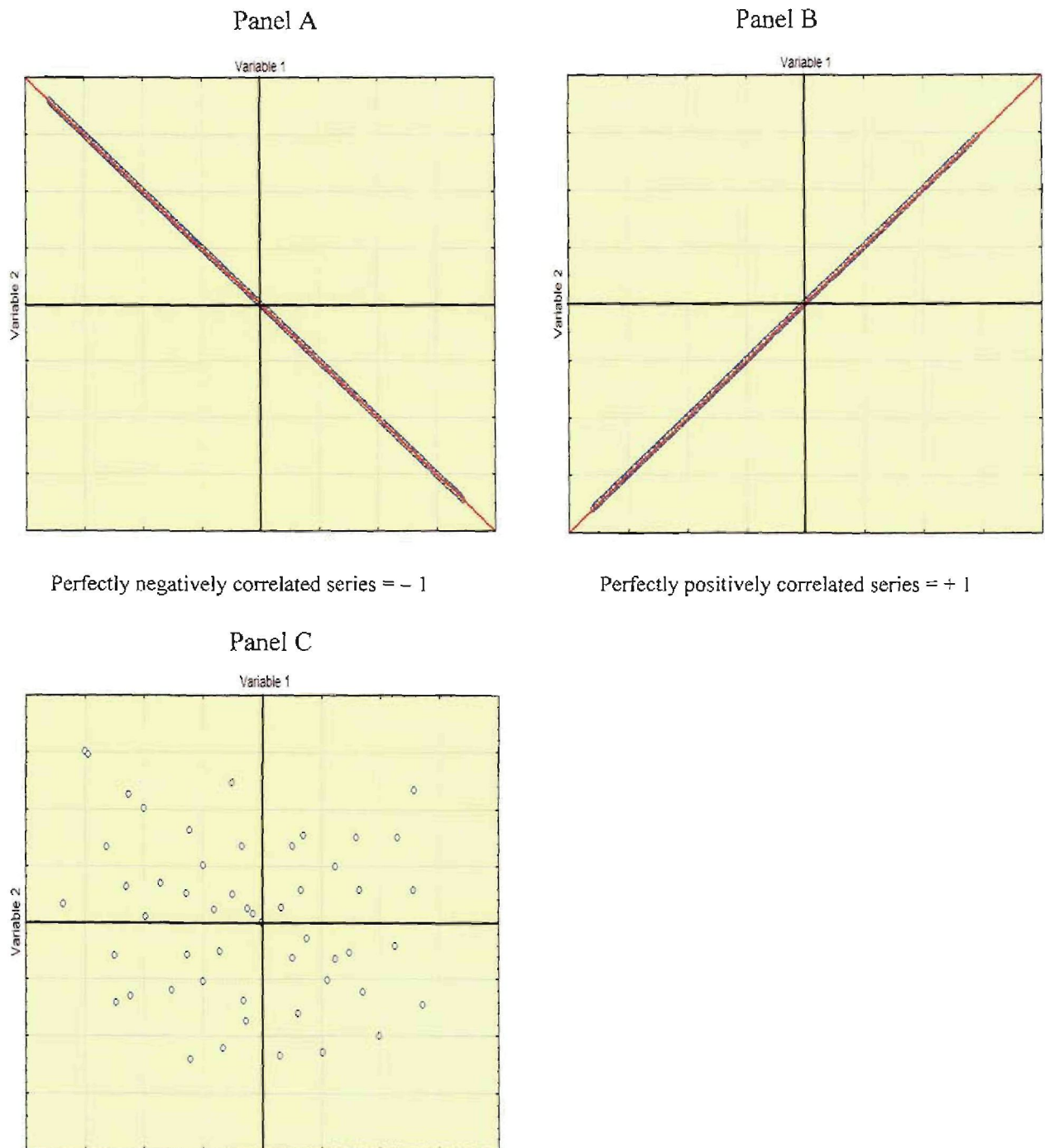
#### 6.2.2.3 Pearson Correlation

The correlation between two variables (whether shares or commodities) reflects the degree to which the variables relate to one another (Koutsoyiannis, 1977:72). Pearson's product-moment correlation coefficient ( $\rho$ ), also referred to as the sample correlation coefficient, is the most common measure of correlation. The correlation coefficient ( $\rho$ ) is a measure of the extent to which two asset samples' rates of return are linearly related (Levy, 2002:180). Pearson's correlation can assume any value between  $-1$  and  $+1$ , depending upon the degree of the relationship. This is illustrated in figure 6.1 below.

---

<sup>73</sup> Please note that it is only the original series that will undergo tests for Pearson and Spearman rank.

FIGURE 6.1 CORRELATION ANALYSIS



Perfectly negatively correlated series =  $-1$

Perfectly positively correlated series =  $+1$

Uncorrelated series =  $0$

(Hirt and Block, 1993:683)

The values  $-1$  and  $+1$  indicate perfect negative and positive relationships respectively. A value of zero indicates that the variables do not co-vary linearly. The degree of co-movement between return variables is particularly important for risk measurement, hedging, and asset allocation (DeFusco, Mcleavy, Pinto & Runkle, 2007a:499).

Pearson's correlation is typically given by the formula (Huntington, 1919:426):

$$r = \frac{\frac{1}{n} \sum (xy) - \bar{x}\bar{y}}{\sqrt{\frac{1}{n} \sum (x^2) - \bar{x}^2} \sqrt{\frac{1}{n} \sum (y^2) - \bar{y}^2}}, \quad (6.19)$$

where  $r$  is the correlation coefficient, and where in the nominator:

$$\sigma_x = \sqrt{\frac{1}{n} \sum (x^2) - \bar{x}^2},$$

and

$$\sigma_y = \sqrt{\frac{1}{n} \sum (y^2) - \bar{y}^2}$$

are the standard deviations of the  $x$ 's and  $y$ 's. According to Hull (2000:382), correlation can also be given by:

$$\rho_{A,B} = \frac{Cov(R_A, R_B)}{\sigma_A \sigma_B}, \quad (6.20)$$

where  $\rho_{A,B}$  is the correlation between the rates of return between assets A and B,  $Cov(R_A, R_B)$  is the covariance between assets A and B, and  $\sigma_A \sigma_B$  are the volatilities of assets A and B.

Although covariance is not discussed in this section, it is to be noted that covariance can also be used as a measure of the relationship between two variables (Levy, 2002:180). Correlation is a superior measure of dependency between variables than covariance, in that a greater numeric value of correlation indicates a stronger dependency between the variables.

Pearson correlation is a well-known tool used to indicate whether variables have a positive or negative statistical relationship with each other. In the case where the

Pearson correlation coefficient is close to zero, it is an indication that there is no statistical relationship between the variables over time.

Pearson's correlation was used in both the long- and short-run for the purpose of assessing the statistical relationship between agricultural commodity series. In section 6.2.2.5 below, the testing for Pearson correlation, as used to test for the statistical relationships between the various long-run pairs selected, is discussed in more detail. This was done by acquiring the Pearson correlation coefficient for each of the long-run pairs over the full period of the data as part of the pairs selection process.

#### 6.2.2.4 Utilising Pearson's Correlation to Validate Long-run Pairs

##### 6.2.2.4.1 Introduction

The literature suggests two methods when making use of correlation as a tool in the pair selection process.<sup>74</sup> The first entails de-trending the data by taking the first difference in the log prices of the variables to ensure stationarity, thus leaving all inference to be based on returns (Alexander & Dimitriu, 2002:3). Once the data have been de-trended, the correlation coefficient can be used to find a relationship between the two variables.

The second way of using correlation is by measuring the correlation of the residuals of the price against a 'best-fitting' line drawn through the price series. This approach allows for the measurement of the association of the mean reversion process between the two maize series. Once the correlation coefficient of each pair has been calculated, the feasibility of each pair can be determined by using the strength of the relationship between the two series as an indication. Most pairs traders agree that a correlation of 0.7 or above should be used to ensure that the two commodities have a consistent relationship (Preston, 2005:41).

##### 6.2.2.4.2 Testing for Pearson's Correlation: Methodology

Pearson's correlation can be applied in more than one way to assess the validity of a pair. In the simplest example, the correlation between the prices of the commodities can

---

<sup>74</sup> It is to be noted that Pearson's correlation was also used in deciding when to execute trades. This use of Pearson's correlation, therefore, is not discussed as part of the pairs selection process.

be measured. Even though this method is used as is by some traders, it can be improved by taking the daily natural logarithmic changes in the prices and by incorporating the measure of mean reversion between the two series (McEwan, 2003:34). This improvement on finding the correlation between two price series is the log returns method, which measures the correlation between the daily natural logarithmic changes (daily return) in their prices given by Castleman (2003:39):

$$R_t = \ln \left( \frac{P_t}{P_{t-1}} \right), \quad (6.21)$$

where  $R_t$  is today's return,  $P_t$  is today's closing price, and  $P_{t-1}$  is yesterday's closing price.

Although it is better to measure the correlation between the logarithmic changes in prices, this method still ignores mean reversion. To get a better idea of the mean reversion process between the two commodities, the correlation of the residuals of price against a regression line (best-fitting line) drawn through the price series can be measured (McEwan, 2003:35). This method of measurement is preferred over the log returns method by McEwan (2003:34 & 35). Many econometric software packages, such as, SAS<sup>®</sup> (SAS, 2007a) and EViews<sup>®</sup> (QMS, 2004b) make provision for fitting a best-fitting straight line through the data using the OLS method. If the original data series is given by:

$$y_t = \beta_0 + \beta_1 X_t, \quad (6.22)$$

where  $\beta_0$  is the intercept and  $\beta_1$  the slope, the best-fitting line through the data can be given by:

$$\hat{y}_t = b_0 + b_1 x_t, \quad (6.23)$$

where  $x = X$ ,  $b_0 = \beta_0$ , and  $b_1 = \beta_1$ . In order to calculate the residual values for each data point, the estimated values of  $y$  are subtracted from original  $y$  values. This residuals equation takes the form of:

$$\varepsilon_t = y_t - \hat{y}_t. \quad (6.24)$$

In order to cross-check for the co-integration results, both the returns and mean reversion methods were tested over the full period of the data for each pair.<sup>75</sup> Pearson's correlation was employed in two ways: taking the correlation of the natural logarithmic returns of the long-run pairs, and taking the correlation of the residuals from the original price series against a regression line drawn through the price series over the full period of the data.

#### 6.2.2.4.3 Testing for Pearson's Correlation: Results

Once a pairs portfolio based on the long-run relationship of the different pairs has been assembled, it is imperative that the trader monitors these relationships as they must continue to hold if he is to be successful. The first test performed indicated the relationship of the various long-run pairs based on the log return prices only. These results are displayed in Table 6.6. The second test indicated the correlation of the two series' residuals based on the residuals data; the results of this test are reported in Table 6.7.

TABLE 6.6 PEARSON'S TEST FOR CORRELATION: LOG RETURNS

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.827	0.786	0.740	0.729	0.716	0.653	0.603
WM 1 away	0.827	1.000	0.897	0.805	0.650	0.796	0.724	0.633
WM 2 away	0.786	0.897	1.000	0.878	0.625	0.774	0.793	0.685
WM 3 away	0.740	0.805	0.878	1.000	0.584	0.692	0.712	0.752
YMSPOT	0.729	0.650	0.625	0.584	1.000	0.760	0.673	0.619
YM 1 away	0.716	0.796	0.774	0.692	0.760	1.000	0.840	0.744
YM 2 away	0.653	0.724	0.793	0.712	0.673	0.840	1.000	0.685
YM 3 away	0.603	0.633	0.685	0.752	0.619	0.744	0.685	1.000

<sup>75</sup> In the description of the results for Pearson correlation, Spearman rank correlation as well as Beta reference is made to the log returns data and the residuals data. This is done to indicate the difference in the methods used to test for the statistical relationship between the commodities that make up each pair.

TABLE 6.7 PEARSON'S TEST FOR CORRELATION: RESIDUALS

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.990	0.975	0.973	0.952	0.946	0.943	0.946
WM 1 away	0.990	1.000	0.992	0.986	0.953	0.964	0.966	0.965
WM 2 away	0.975	0.992	1.000	0.995	0.936	0.959	0.974	0.974
WM 3 away	0.973	0.986	0.995	1.000	0.929	0.947	0.964	0.974
YMSPOT	0.952	0.953	0.936	0.929	1.000	0.985	0.966	0.960
YM 1 away	0.946	0.964	0.959	0.947	0.985	1.000	0.992	0.980
YM 2 away	0.943	0.966	0.974	0.964	0.966	0.992	1.000	0.974
YM 3 away	0.946	0.965	0.974	0.974	0.960	0.980	0.974	1.000

According to the logarithmic returns results in Table 6.6, only seventeen of the twenty-eight possible long-run pairs have a Pearson correlation coefficient of 0.7 and higher, while only five long-run pairs have a Pearson correlation coefficient of 0.8 and higher. The relationship between WM 1 away & WM 2 away is the strongest (0.897), followed by WM 2 away & WM 3 away (0.878), and then YM 1 away & YM 2 away (0.840). The weakest relationship exists between YMSPOT & WM 3 away (0.584), followed by WMSPOT & YM 3 away (0.603), and then YMSPOT & YM 3 away (0.619).

The mean reversion results are more promising, indicating that all the long-run pairs have Pearson correlation coefficients of 0.9 and higher. The strongest relationship was found between WM 2 away & WM 3 away (0.995), followed by WM 1 away & WM 2 away as well as YM 1 away & YM 2 away (both 0.992), and then WM 1 away & WM 3 away (0.986). The relationship between YMSPOT & WM 3 away is the weakest (0.927), followed by YMSPOT & WM 2 away (0.935), and then WMSPOT & YM 2 away (0.943).

#### 6.2.2.4.4 Conclusion

Since the residual data allows for the mean reversion method, this method was followed in deciding on the pairs that were to be included in the final portfolio.<sup>76</sup>

<sup>76</sup> According to the literature, the mean reversion method gives a better indication of the statistical relationship between the series.

From the results reported above it is clear that all the long-run pairs have a stable statistical relationship over the period of the data (with regard to the mean reversion process). This implies that although it is possible for this statistical relationship to fade for short periods of time, the relationship will always be restored in the longer-run. It is this long-run relationship that will allow the pairs trader to trade the two commodities successfully as a pair.

Since each of the long-run pairs has a Pearson correlation coefficient of 0.7 and higher, the statistical relationship of each pair could be further validated by making use of Spearman's rank correlation coefficient. Section 6.2.2.5 discusses the validation of the long-run pairs' relationships by making use of Spearman's rank correlation coefficient.

#### 6.2.2.5 Spearman Rank Correlation

Spearman's rank correlation coefficient is largely equivalent to Pearson's correlation discussed above. Contrary to Pearson's correlation coefficient, however, the Spearman rank correlation coefficient ( $r_s$ ) is a non-parametric measure of correlation (Spearman, 1904:81). Unlike the Pearson correlation coefficient, it does not assume that the relationship between the variables is linear, nor does it require the variables to be measured on interval scales.

In principle, Spearman's rank correlation coefficient is simply a special case of the Pearson product-moment coefficient in which the data are converted to rankings before calculating the coefficient (Spearman, 1904). In practice, however, a simpler procedure is normally used to calculate  $r_s$ . The raw scores are converted to ranks, and the differences between the ranks of each observation on the two variables are calculated. Spearman's rank correlation coefficient is then given by (DeFusco *et al.*, 2007b:451):

$$r_s = \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (6.25)$$

where  $d_i$  is the difference in rank for the  $i$ -th observation of both variables in the different series, and  $n$  is the total number of observations.

Spearman's rank correlation was used in the same way Pearson's correlation was used to find the long-run correlation and the correlation coefficients of the two variables' residuals.

### 6.2.2.6 Utilising Spearman Rank Correlation to Validate Long-run Pairs

#### 6.2.2.6.1 Introduction

In order to test for Spearman's rank correlation, observations are ranked from the largest to the smallest values before the correlation between the ranks are tested. The values in the selected sample can be ranked from a maximum to a minimum value or from a minimum to a maximum value. In testing for the correlation between commodities, the lowest rank was assigned to the smallest value in the sample, and the largest rank to the largest value in the sample.

As with Pearson's correlation, Spearman's correlation coefficient was used to indicate the strength of the statistical relationship between two commodities over the long-run. This was done by testing the correlation over the full period of the data. Commodities with high coefficient values were considered pairs, and the correlation of the residuals of price against a regression line drawn through the price series was used to incorporate the mean reversion process between the two commodities.

#### 6.2.2.6.2 Testing for Spearman's Rank Correlation: Methodology

Since both the co-integration and Pearson correlation tests indicated that all the pairs do indeed have strong statistical relationships, the Spearman rank correlation test was performed next. In order to calculate the correlation between the two data series, each of the data points (returns values) were ranked out of 2029. The correlation between these ranked values was then calculated.

Again, tests were performed by making use of two methods: firstly, using the correlation coefficients for each long-run pair based on the log returns data; and secondly, using the correlation coefficients for each long-run pair based on the residuals

data.<sup>77</sup> As with the Pearson correlation tests, both the returns and mean reversion methods were tested over the full duration of the data period for each pair. Spearman's rank correlation was thus used in two ways: in measuring the correlation of the logarithmic returns of each of the long-run pairs over the full period of the data, and in measuring the correlation of the residuals from the original price series against a regression line drawn through the price series over the full period of the data for each pair.

#### 6.2.2.6.3 Testing for Spearman's Rank Correlation: Results

Spearman's rank correlation coefficient is a non-parametric measure of correlation that allows the pairs trader to ascertain whether the long-run relationship between the two commodities will hold. As with the Pearson correlation tests above, the first test was performed to establish the relationship of the various long-run pairs based on the log return prices, while the second test was performed to establish what the correlation coefficient of the two series' residuals is, based on the residuals data. The results of the first test are given in Table 6.8 below, and those of the second test, in Table 6.9.

TABLE 6.8 SPEARMAN'S RANK TEST FOR CORRELATION: LOG RETURNS

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.867	0.824	0.784	0.711	0.719	0.681	0.634
WM 1 away	0.867	1.000	0.923	0.870	0.661	0.778	0.739	0.674
WM 2 away	0.824	0.923	1.000	0.904	0.644	0.758	0.767	0.694
WM 3 away	0.784	0.870	0.904	1.000	0.615	0.719	0.719	0.712
YMSPOT	0.711	0.661	0.644	0.615	1.000	0.781	0.718	0.650
YM 1 away	0.719	0.778	0.758	0.719	0.781	1.000	0.853	0.781
YM 2 away	0.681	0.739	0.767	0.719	0.718	0.853	1.000	0.694
YM 3 away	0.634	0.674	0.694	0.712	0.650	0.781	0.694	1.000

<sup>77</sup> See equation 6.21 for a description of the calculation of the natural logarithm of returns, and equation 6.24 for the calculation of the residual values.

TABLE 6.9 SPEARMAN'S RANK TEST FOR CORRELATION: RESIDUALS

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.974	0.948	0.936	0.938	0.911	0.896	0.890
WM 1 away	0.974	1.000	0.988	0.970	0.921	0.941	0.941	0.926
WM 2 away	0.948	0.988	1.000	0.988	0.891	0.927	0.953	0.946
WM 3 away	0.936	0.970	0.988	1.000	0.877	0.904	0.939	0.956
YMSPOT	0.938	0.921	0.891	0.877	1.000	0.969	0.935	0.919
YM 1 away	0.911	0.941	0.927	0.904	0.969	1.000	0.980	0.951
YM 2 away	0.896	0.941	0.953	0.939	0.935	0.980	1.000	0.946
YM 3 away	0.890	0.926	0.946	0.956	0.919	0.951	0.946	1.000

The results of the log returns test in Table 6.8 indicate that nineteen of the twenty-eight possible long-run pairs have a Spearman rank correlation coefficient of 0.7 and higher, while only six long-run pairs have a Spearman rank correlation coefficient of 0.8 and higher. The strongest statistical relationship was found between WM 1 away & WM 2 away (0.923), followed by WM 2 away & WM 3 away (0.904), and then WM 1 away & WM 3 away (0.870). The weakest relationship exists between YMSPOT & WM 3 away (0.615), followed by WMSPOT & YM 3 away (0.634), and then YMSPOT & WM 2 away (0.644).

As with the Pearson correlation results, the correlation coefficients for the log returns tests are lower than that of the residuals tests. According to the mean reversion (residuals) results, all the long-run pairs have Spearman correlation coefficients of 0.8 and higher, while twenty-four of the twenty-eight possible long-run pairs have a Spearman rank correlation coefficient of 0.9 and higher. The relationship between WM 1 away & WM 2 away as well as WM 2 away & WM 3 away is the strongest (both 0.988), followed by YM 1 away & YM 2 away (0.980). The weakest relationship exists between YMSPOT & WM 3 away (0.877), followed by WMSPOT & YM 3 away (0.890), and then YMSPOT & WM 2 away (0.891).

#### 6.2.2.6.4 Conclusion

Although the logarithmic returns results indicate that only nineteen of the twenty-eight possible long-run pairs have a Spearman rank correlation coefficient of 0.7 and higher, the mean reversion results report correlation coefficients of 0.8 and higher for all the

long-run pairs. From these results is clear that all the long-run pairs have a stable statistical relationship over the data period.

Again, as with the Pearson correlation tests, the mean reversion results were followed in deciding on the pairs that were to be included in the final portfolio. Since each of the long-run pairs has a Pearson correlation coefficient of 0.7 and higher, the statistical relationship of each pair underwent further validation by making use of the seasonal measures described in section 6.4. The next section discusses Beta and its role in trading pairs.

### 6.2.2.7 Beta

Beta indicates the volatility or risk of a fund in comparison to that of its index or benchmark (FTSE, 2005:16). A fund with a Beta very close to one indicates that the fund's performance closely matches the index or benchmark. A Beta greater than one indicates greater volatility than the overall market volatility, while a Beta less than one indicates less volatility than the benchmark (Croome, 2003:2). A greater Beta value is thus an indication of higher market (non-diversifiable) risk.

Beta can also be used to indicate the relationship between two variables, such as individual shares. This relationship between variables can be positive or negative: a positive Beta value suggests that both the market's return and the asset's return move in the same direction when market conditions change. Even though correlation and Beta may appear similar, the difference between the two is that the magnitude of the Beta value is not an indication of the strength of the relationship between the two variables. Rather, the magnitude of the value of Beta is an indication of the effect that an external shock will have on that variable given the reaction of the benchmark variable on the shock. In order to ensure the full benefit of its use it is vital that Beta be used in conjunction with other statistical measurement tools, such as co-integration and correlation.

Although Beta is usually used to measure the risk, this tool might be helpful in deciding when to enter or exit a pairs trade. Since Beta can provide an indication of how closely a price series matches the performance of an index or another benchmark, this tool can be

utilised by the pairs trader to ascertain how closely the performance of one maize series matches that of another. Beta is also a useful tool to use in order to establish how volatile one maize series is when measured against another. The next section continues discussing the use of Beta to confirm a long-run pair.

### 6.2.2.8 Utilising Beta to Validate Long-run Pairs

#### 6.2.2.8.1 Introduction

Although Beta is usually employed to determine the volatility of a fund in comparison to that of its index or benchmark (FTSE, 2005:16), this study made use of Beta as a measure of the statistical relationship between two commodities. Since Beta expresses the movement of one variable based on a 1 percent change in another, it can be used to indicate the co-movement between two individual data series. In order to utilise Beta for this purpose, one of the maize series' percentage returns was regressed on the other. One of the maize series' percentage returns was set as the dependent variable, and the other maize series' percentage returns was set as the independent variable. Beta then indicates the magnitude of the relationship of the independent variable relative to the dependent variable (Preston, 2005:41).

By making use of Beta in this manner, it was possible to assess whether there is a strong enough relationship between two commodities to consider them a pair. It should be noted that Beta was only used as an indication of the co-movement of the two commodities in the pair, and not as an indication of the strength of the relationship between the two commodities.

#### 6.2.2.8.2 Testing for Beta: Methodology

Similar to both Pearson's correlation and Spearman's rank correlation, tests for Beta were performed by making use of two methods. The value of Beta was calculated firstly by making use of the daily natural log changes (daily returns) in the price, and secondly, by making use of the residuals data.<sup>78</sup> As was the case with the two correlation tests,

---

<sup>78</sup> See equation 6.21 above for the description of the calculation of the natural logarithm of returns, and equation 6.24 for the calculation of the residual values.

both the returns and mean reversion methods were tested over the full duration of the data period for each pair.

The value of Beta for a data series was calculated using Microsoft® Excel® (MS, 2007). In order to test for Beta coefficients of the long-run pairs selected during the co-integration tests in Excel®, the calculation was divided in two steps. The first step simply required calculating the covariance of the two series, while the second step entailed calculating the variance of the benchmark (independent variable) series. Once the covariance between the two series and the variance of the benchmark series was known, the Beta of the dependent variable was calculated as follows (Gardner & Mills, 1994:144):

$$\beta_i = \frac{\text{cov}(p_i, p_j)}{\sigma_j}, \quad (6.26)$$

where  $p_i$  is the price of commodity  $i$ ,  $p_j$  is the price of commodity  $j$  and  $\sigma_j$  is the variance of commodity  $j$ . The following section reports on the results of the tests performed on all the long-run pairs regarding their Beta values.

#### 6.2.2.8.3 Testing for Beta: Results

The final step in the pairs selection process was verifying the validity of the long-run pairs that were selected during the co-integration process by means of Beta. As was the case with the two correlation tests, the first test for Beta values made use of log return prices to establish the relationship of the various long-run pairs, while the second test established the relationship of the two series' residuals, based on the residuals data. The results of the first test are displayed in Table 6.10 below, and those of the second test, in Table 6.11.

TABLE 6.10 BETA TEST: LOG RETURNS

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.892	0.871	0.838	0.827	0.894	0.773	0.769
WM 1 away	0.892	1.000	0.921	0.845	0.684	0.921	0.795	0.748
WM 2 away	0.871	0.921	1.000	0.897	0.640	0.872	0.847	0.788
WM 3 away	0.838	0.845	0.897	1.000	0.585	0.763	0.744	0.846
YMSPOT	0.827	0.684	0.640	0.585	1.000	0.836	0.702	0.695
YM 1 away	0.894	0.921	0.872	0.763	0.836	1.000	0.796	0.760
YM 2 away	0.773	0.795	0.847	0.744	0.702	0.796	1.000	0.788
YM 3 away	0.769	0.748	0.788	0.846	0.695	0.760	0.788	1.000

TABLE 6.11 BETA TEST: RESIDUALS

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.896	0.845	0.791	0.980	0.930	0.882	0.842
WM 1 away	0.896	1.000	0.943	0.883	1.094	1.038	0.985	0.940
WM 2 away	0.845	0.943	1.000	0.937	1.160	1.100	1.044	0.997
WM 3 away	0.791	0.883	0.937	1.000	1.238	1.174	1.115	1.064
YMSPOT	0.980	1.094	1.160	1.238	1.000	0.949	0.901	0.860
YM 1 away	0.930	1.038	1.100	1.174	0.949	1.000	0.949	0.906
YM 2 away	0.882	0.985	1.044	1.115	0.901	0.949	1.000	0.997
YM 3 away	0.842	0.940	0.997	1.064	0.860	0.906	0.997	1.000

Testing for the Beta values of the long-run pairs selected previously, revealed a positive relationship between each of the commodities that constitutes every pair. This result indicates that each of the maize series in every pair moves in the same direction over time.

The Beta values of the residuals tests are once again higher than those of the tests performed using the logarithmic returns method. The results of the logarithmic returns test in Table 6.10 indicates that twenty-four of the twenty-eight possible long-run pairs have Beta values of 0.7 and higher, while thirteen long-run pairs have Beta values of 0.8 and higher. None of the pairs have Beta values reaching 1, indicating that each movement of the logarithmic returns of the benchmark maize series caused the

logarithmic returns of the dependent maize series to move in the same direction (positive Beta), but to a lesser extent.<sup>79</sup>

The pair with the highest Beta value is 0.921 for the WM 1 away & WM 2 away pair and the WM 1 away & YM 1 away pair, followed by the WM 2 away & WM 3 away pair (0.897). The pair with the lowest Beta value is YMSPOT & WM 3 away (0.585), followed by YMSPOT & WM 2 away (0.640), and then YMSPOT & WM 1 away (0.684).

The results of both Pearson as well as Spearman's rank correlation corroborate the Beta results regarding the YMSPOT combinations. While the Beta values for these combinations indicate that the log returns of the dependent commodities move in the same direction (positive Beta), but to a lesser extent, the correlation results reveal correlation coefficients lower than 0.7 for the same commodity combinations.

The residuals test results in Table 6.11 indicate that nine of the twenty-eight possible long-run pairs have Beta values of 1 and higher, while twenty-seven long-run pairs have Beta values of 0.8 and higher. The pair with the highest Beta value is 1.238 for the YMSPOT & WM 3 away pair, followed by YM 1 away & WM 3 away (1.174) and then YMSPOT & WM 2 away (1.160). That the YSPOT & WM 3 away is the pair with the highest Beta value is curious as this pair was noted as the pair with the weakest statistical relationship by both the Pearson correlation and Spearman rank correlation coefficient tests. The pair with the lowest Beta value is WMSPOT & WM 3 away (0.791), followed by WMSPOT & YM 3 away (0.842) and WMSPOT & WM 2 away (0.845).

#### 6.2.2.8.4 Conclusion

The strength of the relationship of each pair was established by the tests for co-integration, Pearson correlation and Spearman rank correlation. Beta, however, is not an indication of the strength of this relationship but rather is an indication of the statistical relationship between the two maize series in each pair. Therefore, the results of Beta

---

<sup>79</sup> It is to be noted that the Beta value is different (but still positive) when the other commodity in the pair was set as the benchmark commodity.

were simply utilised to indicate the movement of the one maize series in the pair based on the movement of the other. The value of Beta is that it indicates the reaction of one maize series based on the reaction of the benchmark maize series to the same set of external events, and should thus be interpreted in light of this.

### 6.2.3 Conclusion

All of the statistical measures (co-integration, Pearson correlation and Spearman rank correlation), indicated that all twenty-eight maize series were to be included in the preliminary pairs portfolio.

The next step was to assess the validity of the pairs in the portfolio by comparing their deterministic seasonal patterns. This was done by comparing the Henderson-curves of the individual maize series, and then by comparing the seasonal factors series of the individual maize series, which is reported in the section that follows.

## 6.3 Pairs Selection: Deterministic Seasonality Measures

### 6.3.1 Introduction

As mentioned previously, making use of seasonality tests is a new approach to the pairs selection process. Since maize and other agricultural commodities often contain seasonal patterns, it should be possible to select agricultural commodity pairs, based not only on the statistical relationship between them, but also on the co-movement of the underlying seasonal patterns in the data. Table 6.12 presents the combinations of maize series deemed as pairs by the statistical tests that have been reported in section 6.2.

TABLE 6.12 PRELIMINARY PAIRS PORTFOLIO

WMSPOT & WM 1 away	WMSPOT & WM 2 away	WMSPOT & WM 3 away	WMSPOT & YMSPOT
WMSPOT & YM 1 away	WMSPOT & YM 2 away	WMSPOT & YM 3 away	WM 1 away & WM 2 away
WM 1 away & WM 3 away	WM 1 away & YMSPOT	WM 1 away & YM 1 away	WM 1 away & YM 2 away
WM 1 away & YM 3 away	WM 2 away & WM 3 away	WM 2 away & YMSPOT	WM 2 away & YM 1 away
WM 2 away & YM 2 away	WM 2 away & YM 3 away	WM 3 away & YMSPOT	WM 3 away & YM 1 away
WM 3 away & YM 2 away	WM 3 away & YM 3 away	YMSPOT & YM 1 away	YMSPOT & YM 2 away
YMSPOT & YM 3 away	YM 1 away & YM 2 away	YM 1 away & YM 3 away	YM 2 away & YM 3 away

These combinations of maize series were tested for their validity as pairs by making use of the seasonality tests that have been explained in chapter 5. Section 6.3.2 discusses the process of using the deterministic seasonality test (X12) as a tool for selecting pairs, while section 6.3.3 discusses the process of using stochastic seasonality in selecting pairs.

## 6.3.2 Testing for Pairs with X12

### 6.3.2.1 Introduction

The tests that have been reported in section 5.4.2.6 were performed in order to break the original maize series up into several components. Only two of these components are important to the goals of this thesis: the Henderson-curve-series and the seasonal factors series.<sup>80</sup> The co-movement between the various maize series' seasonal patterns was used to further validate the preliminary pairs shown above. Correlation coefficients between these series were used as an indication of the co-movement between the different Henderson-curves and seasonal factors series. If the correlation coefficient between two Henderson-curves or two seasonal factors series exceeded 0.7, the two underlying maize series were considered a pair.

Section 6.3.2.2 discusses the comparison of the various maize series' Henderson-curves, while section 6.3.2.3 discusses the comparisons of the seasonal factors series of all the maize series.

### 6.3.2.2 The Henderson-curve as Pairs Selection Tool

In making use of the Henderson-curve as pairs selection tool, the correlation coefficients between these series were used as an indication of the co-movement between the different Henderson-curves. The correlation test indicated that all twenty-eight maize pairs have a correlation coefficient of 0.9 and more. The statistical relationship between WM 2 away & WM 3 away is the strongest (0.995), followed by WM 1 away & WM 2 away (0.992), and then YM 1 away & YM 2 away.

---

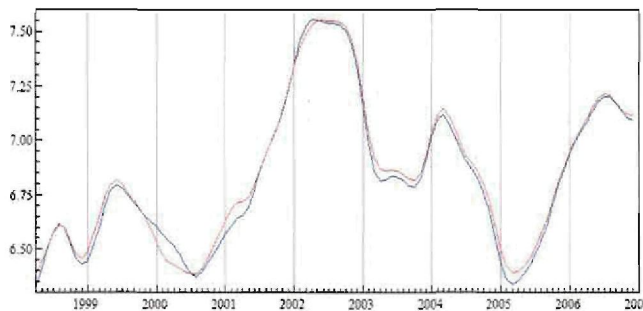
<sup>80</sup> Refer to section 5.4.1 for an explanation of the Henderson-curve and the seasonal factors series.

Conversely, the pairs with the weakest relationships are WM 3 away & YMSPOT (0.929), followed by WM 2 away & YMSPOT (0.936), and then WMSPOT & YM 2 away (0.943). Table 6.13 reports the cross-correlation matrix for the different Henderson-curves, and graphs 6.1 to 6.28 show the comparisons between all the maize series' Henderson-curves.<sup>81</sup>

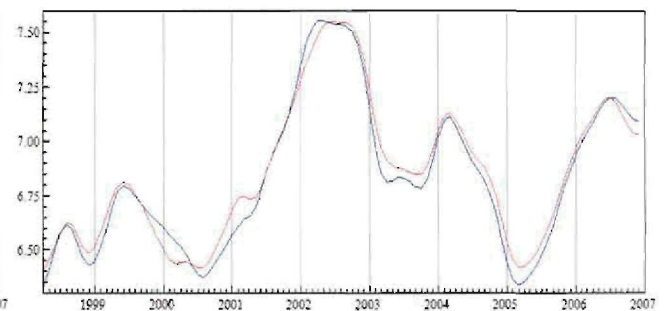
TABLE 6.13 CORRELATION COEFFICIENTS BETWEEN HENDERSON-CURVES

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.990	0.975	0.973	0.952	0.946	0.943	0.946
WM 1 away	0.990	1.000	0.992	0.986	0.953	0.964	0.966	0.965
WM 2 away	0.975	0.992	1.000	0.995	0.936	0.959	0.974	0.974
WM 3 away	0.973	0.986	0.995	1.000	0.929	0.947	0.964	0.974
YMSPOT	0.952	0.953	0.936	0.929	1.000	0.985	0.966	0.960
YM 1 away	0.946	0.964	0.959	0.947	0.985	1.000	0.992	0.980
YM 2 away	0.943	0.966	0.974	0.964	0.966	0.992	1.000	0.974
YM 3 away	0.946	0.965	0.974	0.974	0.960	0.980	0.974	1.000

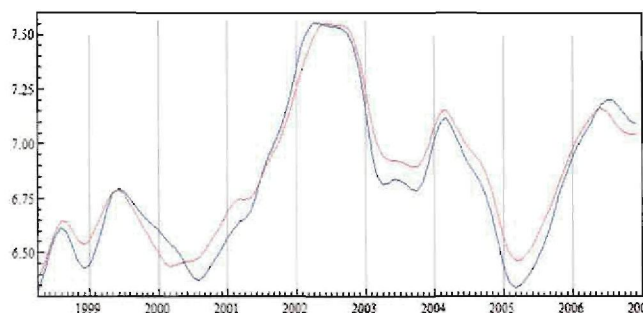
GRAPH 6.1 —WMSPOT —WM 1 away



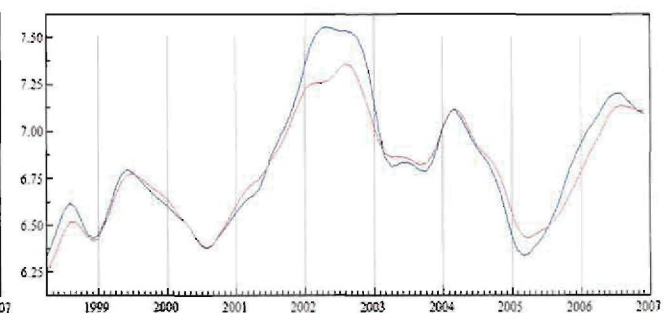
GRAPH 6.2 —WMSPOT —WM 2 away



GRAPH 6.3 —WMSPOT —WM 3 away

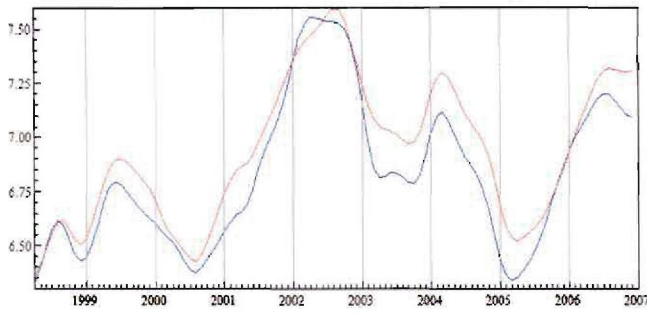


GRAPH 6.4 —WMSPOT —YMSPOT

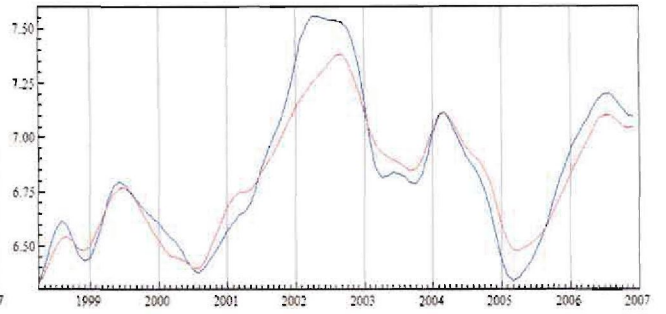


<sup>81</sup> Please note that the Henderson-curve data is given on the y-axis, while the date is given on the x-axis for graphs 6.1 to 6.28.

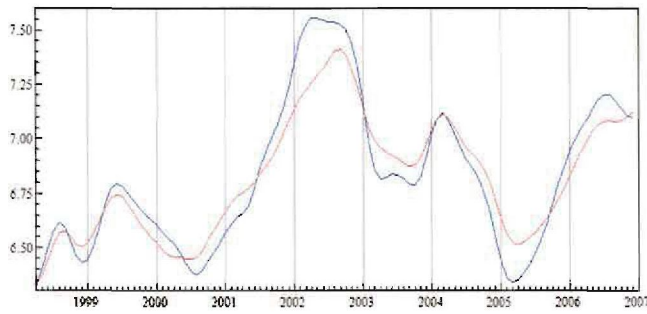
GRAPH 6.5 — WMSPOT — YM 1 away



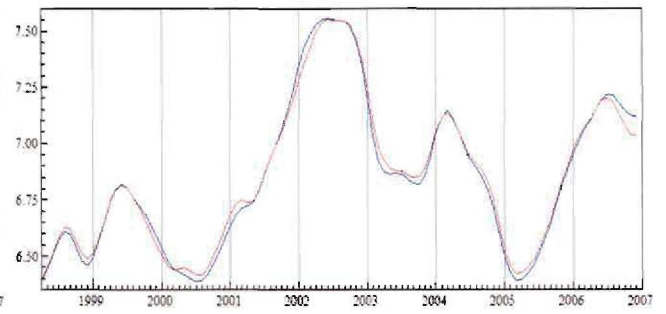
GRAPH 6.6 — WMSPOT — YM 2 away



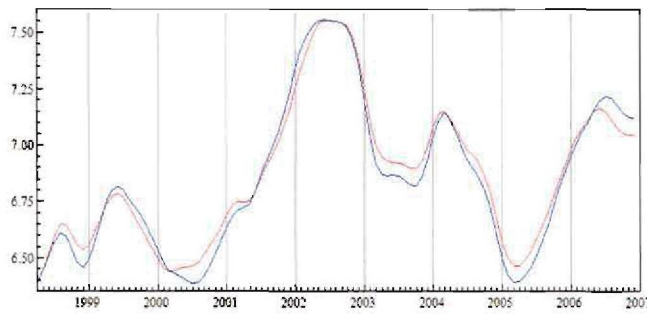
GRAPH 6.7 — WMSPOT — YM 3 away



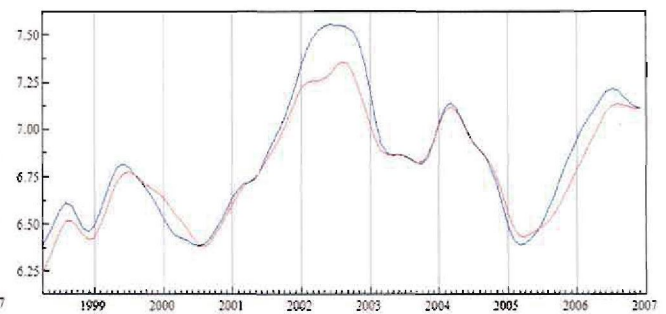
GRAPH 6.8 — WM 1 away — WM 2 away



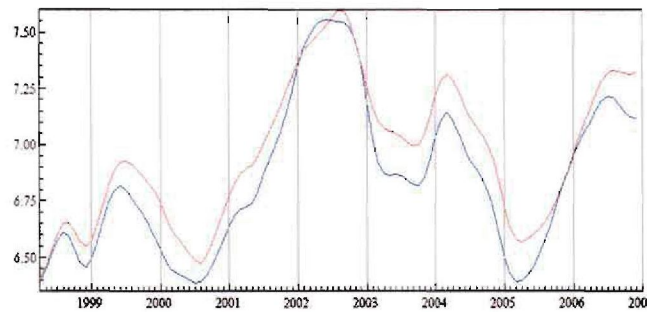
GRAPH 6.9 — WM 1 away — WM 3 away



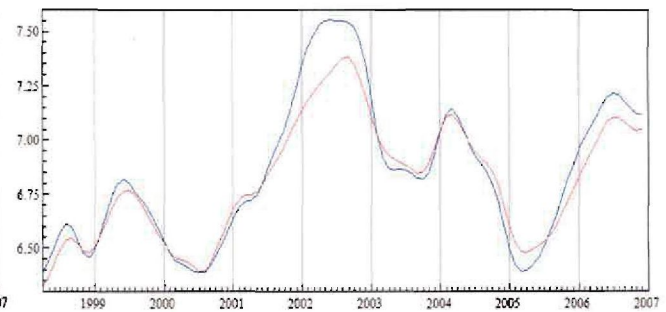
GRAPH 6.10 — WM 1 away — YMSPOT



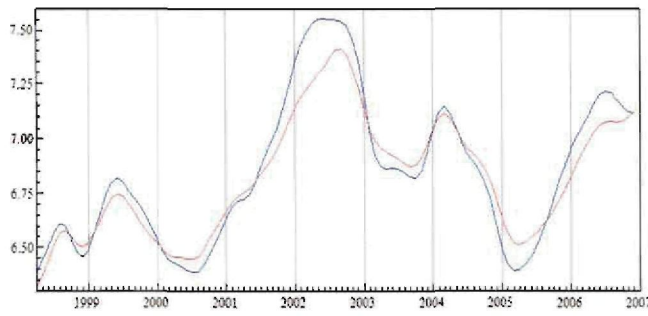
GRAPH 6.11 — WM 1 away — YM 1 away



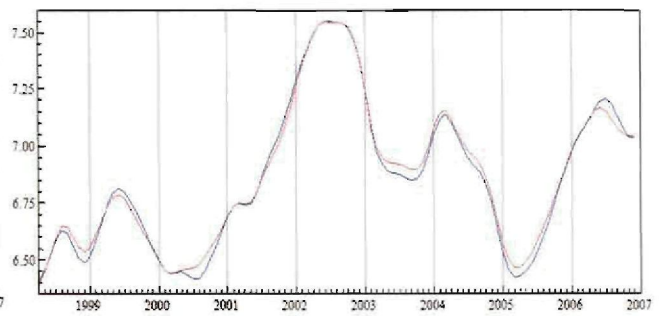
GRAPH 6.12 — WM 1 away — YM 2 away



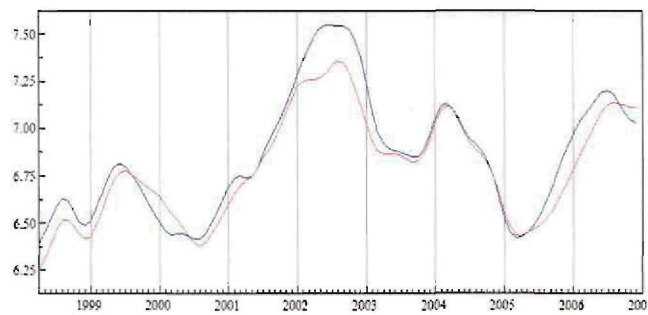
GRAPH 6.13 —WM 1 away —YM 3 away



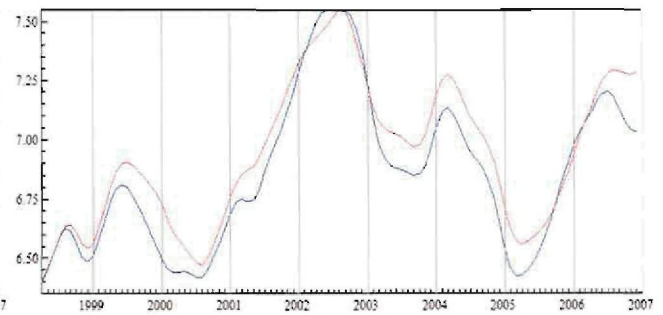
GRAPH 6.14 —WM 2 away —WM 3 away



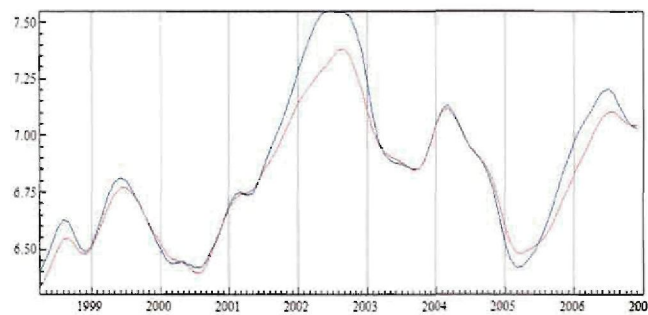
GRAPH 6.15 —WM 2 away —YMSPOT



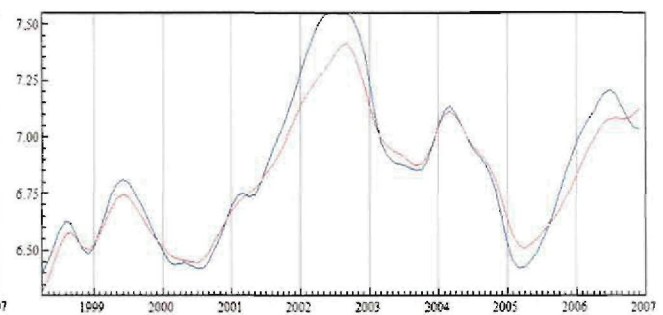
GRAPH 6.16 —WM 2 away —YM 1 away



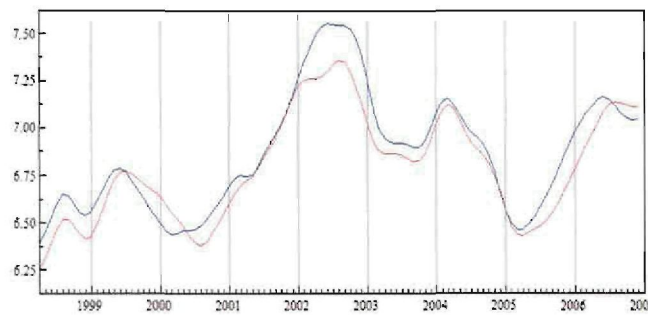
GRAPH 6.17 —WM 2 away —YM 2 away



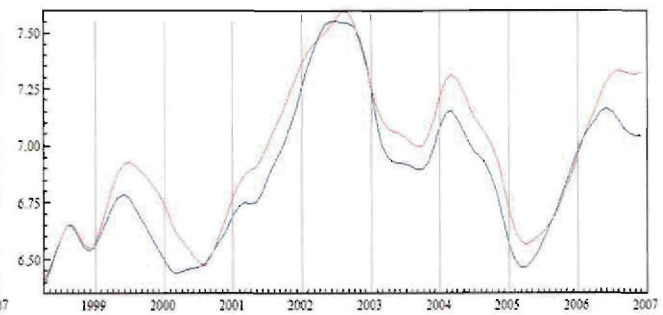
GRAPH 6.18 —WM 2 away —YM 3 away



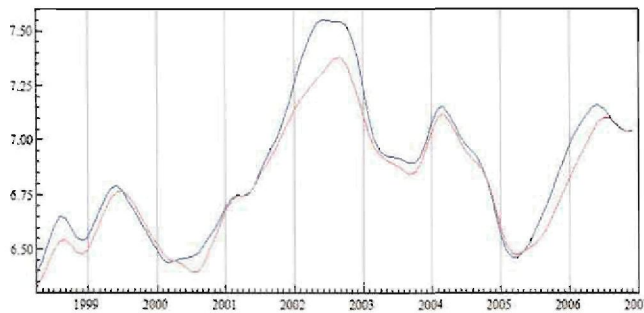
GRAPH 6.19 —WM 3 away —YMSPOT



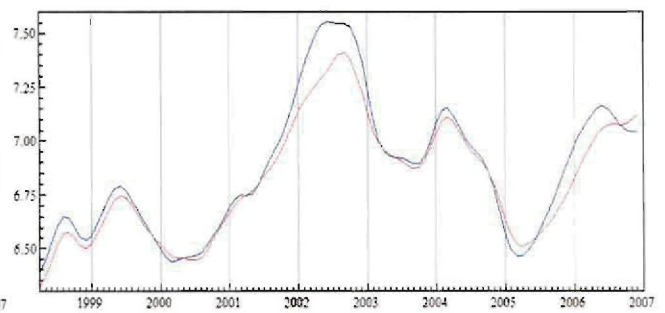
GRAPH 6.20 —WM 3 away —YM 1 away



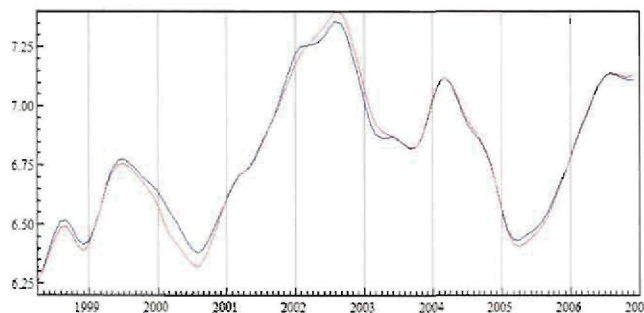
GRAPH 6.21 —WM 3 away —YM 2 away



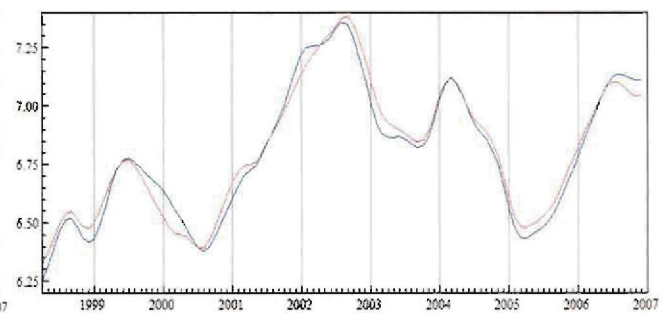
GRAPH 6.22 —WM 3 away —YM 3 away



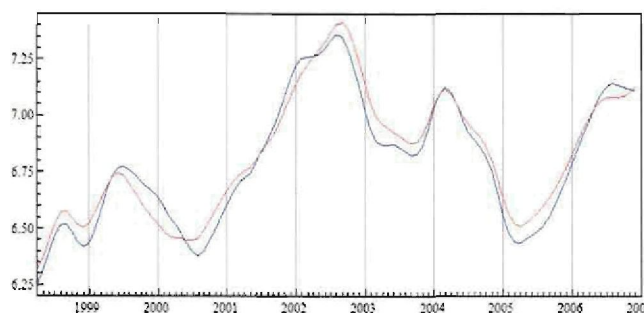
GRAPH 6.23 —YMSPOT —YM 1 away



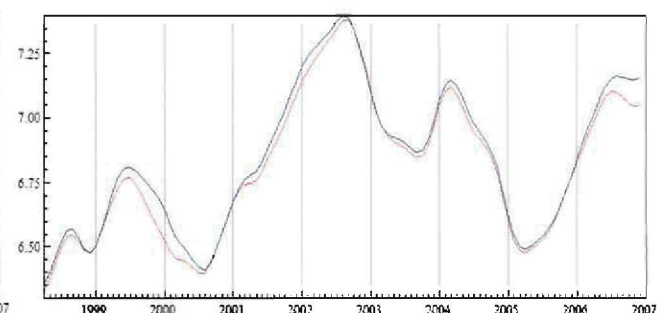
GRAPH 6.24 —YMSPOT —YM 2 away



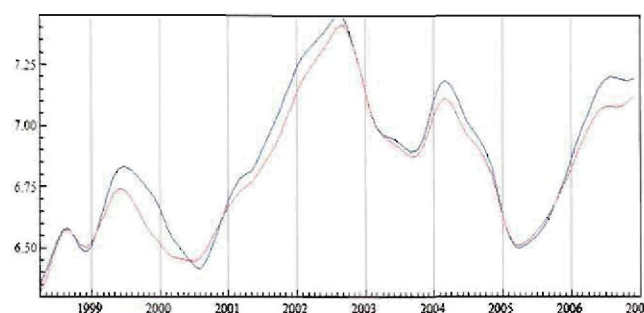
GRAPH 6.25 —YMSPOT —YM 3 away



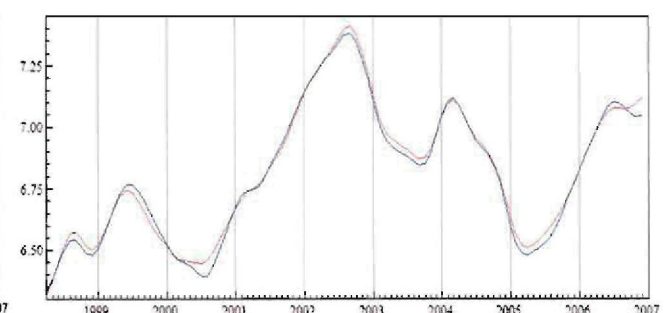
GRAPH 6.26 —YM 1 away —YM 2 away



GRAPH 6.27 —YM 1 away —YM 3 away



GRAPH 6.28 —YM 2 away —YM 3 away



When examining the above graphs, it is clear that all of the preliminary pairs possess similar underlying long-run trends. Further, it is apparent that these long-run trends

differ from one another over time, creating the ideal opportunity for the pairs trader to enter into a pairs transaction. Section 6.3.2.3 further evaluates the validity of the pairs already selected by the statistical measures and the Henderson-curve, using seasonal factors as a pairs selection tool.

### 6.3.2.3 Seasonal Factors as Pairs Selection Tool

The seasonal factors series was employed in the same manner as the Henderson-curve. As before, the correlation coefficients between the seasonal factors series were used as an indication of the co-movement between the different maize series' seasonal patterns. If the correlation coefficient between two maize series was less than 0.7 it was no longer regarded as a valid pair. The correlation test indicated that only seventeen of the twenty-eight maize pairs have a correlation coefficient of 0.7 and more. This means that eleven of the preliminary pairs selected by the previous tests were not included in the final portfolio.

The pair with the strongest statistical relationship is WMSPOT & YMSPOT (0.938), followed by WM 1 away & YM 1 away (0.921), and then WM 3 away & YM 3 away. It is important to remember that this is the correlation coefficient between the two maize series' seasonal factors series not between the actual maize series. Conversely, the pairs with the weakest relationships are YM 1 away & YM 3 away (0.435), followed by WM 3 away & YM 3 away (0.437), and then WM 1 away & YM 3 away (0.517). Table 6.14 reports on the cross-correlation matrix for the different seasonal factors series, and graphs 6.29 to 6.56 show the comparisons between all the maize series' seasonal factors series.<sup>82</sup>

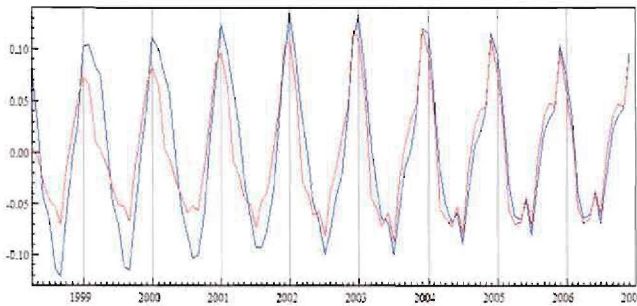
---

<sup>82</sup> Please note that the seasonal factors data is given on the y-axis, while the date is given on the x-axis for graphs 6.29 to 6.56.

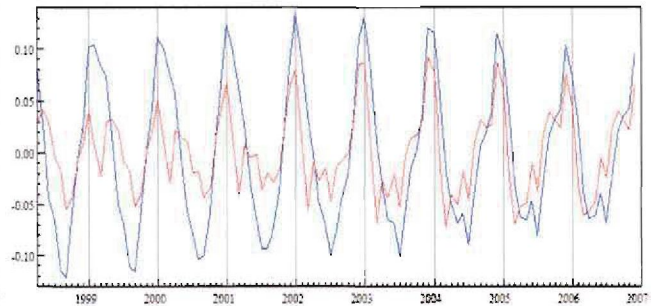
TABLE 6.14 CORRELATION COEFFICIENTS BETWEEN SEASONAL FACTORS SERIES

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.893	0.713	0.754	0.938	0.770	0.661	0.677
WM 1 away	0.893	1.000	0.834	0.631	0.914	0.921	0.841	0.517
WM 2 away	0.713	0.834	1.000	0.814	0.640	0.724	0.900	0.714
WM 3 away	0.754	0.631	0.814	1.000	0.570	0.437	0.602	0.918
YMSPOT	0.938	0.914	0.640	0.570	1.000	0.894	0.701	0.531
YM 1 away	0.770	0.921	0.724	0.437	0.894	1.000	0.874	0.435
YM 2 away	0.661	0.841	0.900	0.602	0.701	0.874	1.000	0.609
YM 3 away	0.677	0.517	0.714	0.918	0.531	0.435	0.609	1.000

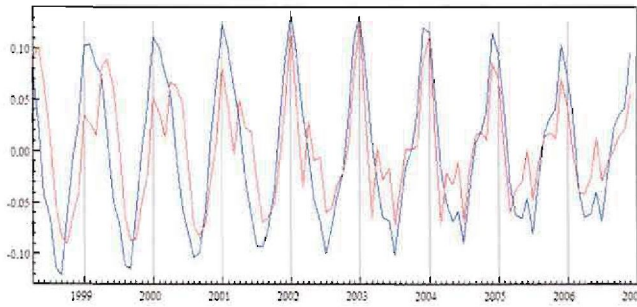
GRAPH 6.29 —WMSPOT —WM 1 away



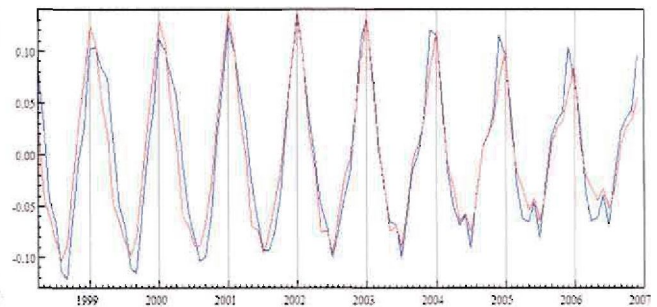
GRAPH 6.30 —WMSPOT —WM 2 away



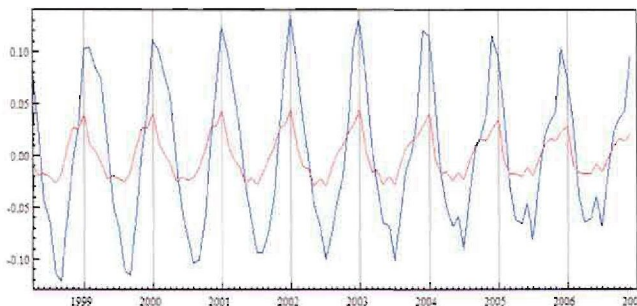
GRAPH 6.31 —WMSPOT —WM 3 away



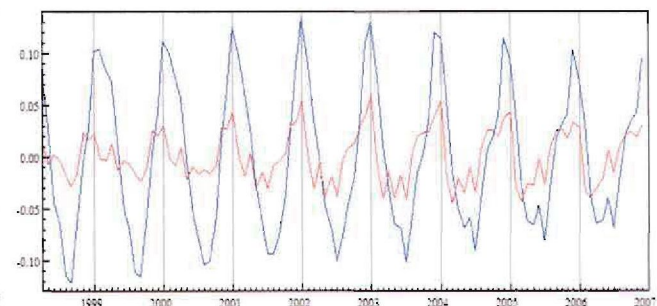
GRAPH 6.32 —WMSPOT —YMSPOT



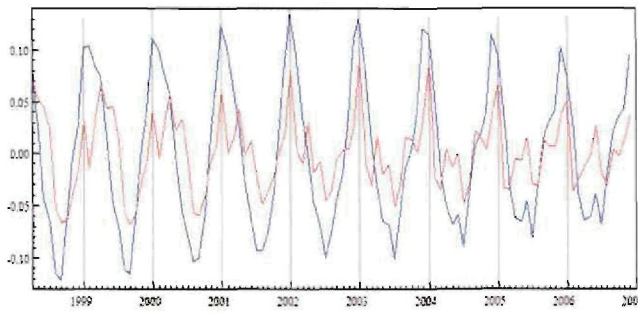
GRAPH 6.33 —WMSPOT —YM 1 away



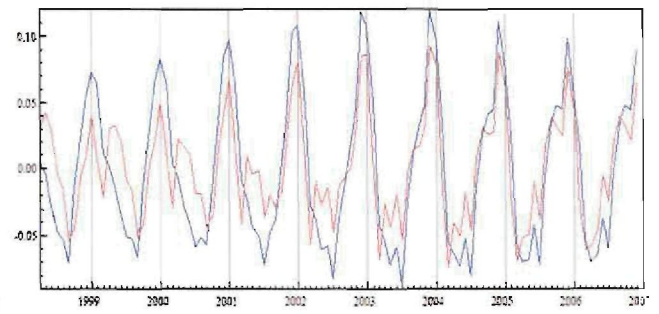
GRAPH 6.34 —WMSPOT —YM 2 away



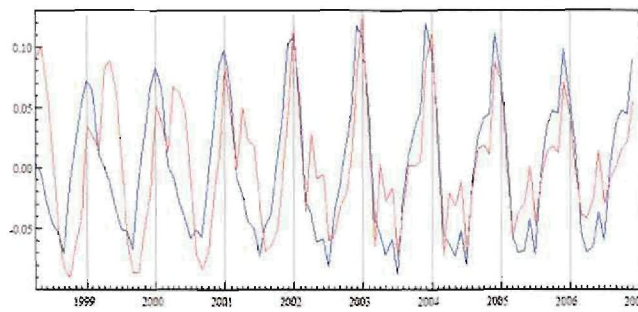
GRAPH 6.35 — WMSPOT — YM 3 away



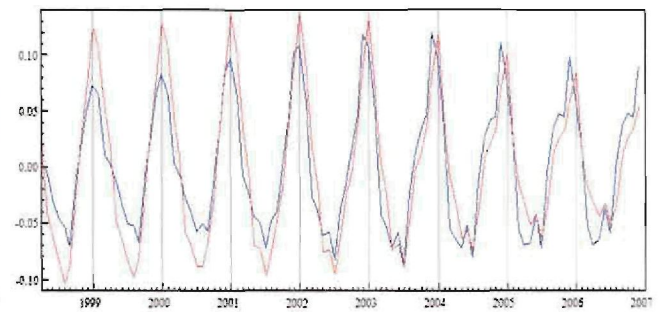
GRAPH 6.36 — WM 1 away — WM 2 away



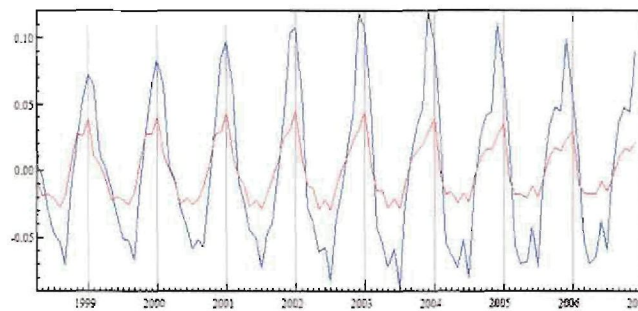
GRAPH 6.37 — WM 1 away — WM 3 away



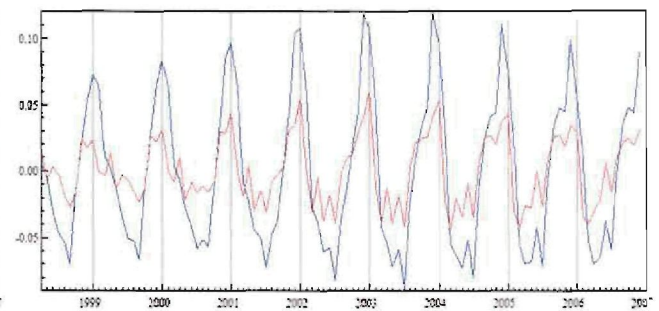
GRAPH 6.38 — WM 1 away — YMSPOT



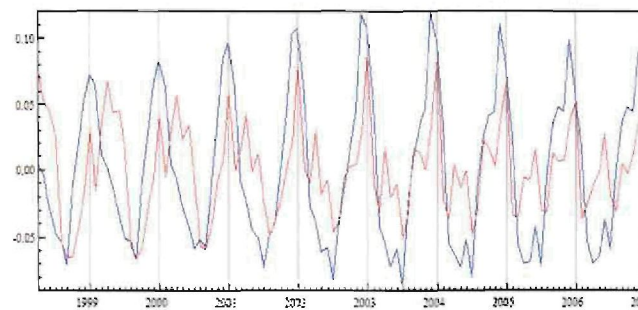
GRAPH 6.39 — WM 1 away — YM 1 away



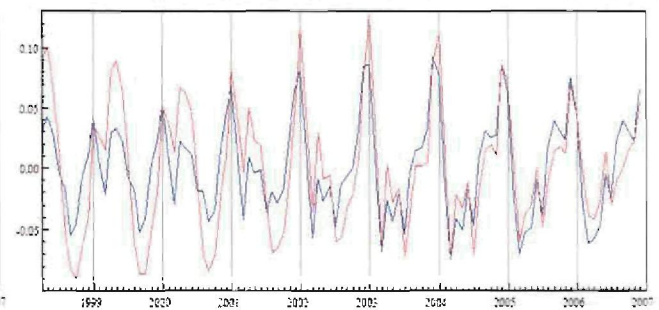
GRAPH 6.40 — WM 1 away — YM 2 away



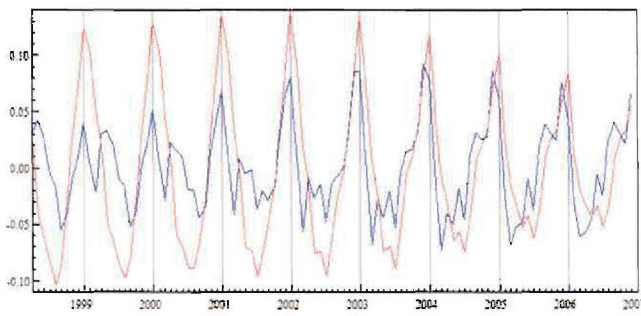
GRAPH 6.41 — WM 1 away — YM 3 away



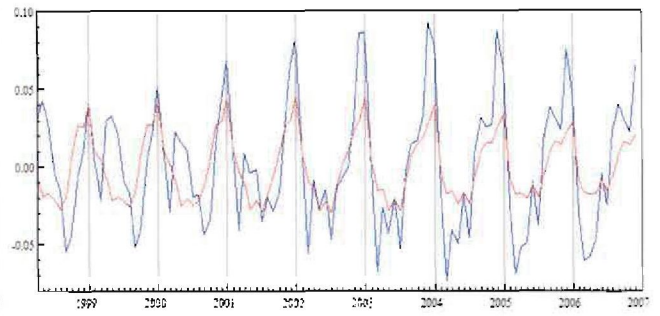
GRAPH 6.42 — WM 2 away — WM 3 away



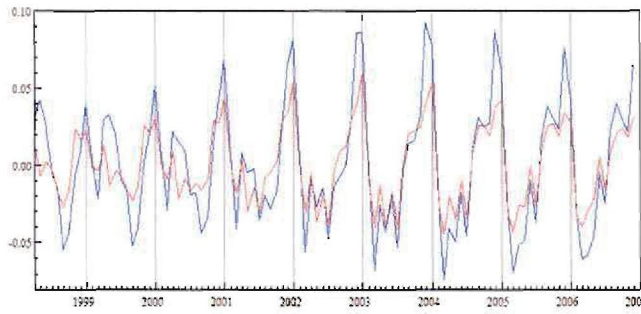
GRAPH 6.43 —WM 2 away —YMSPOT



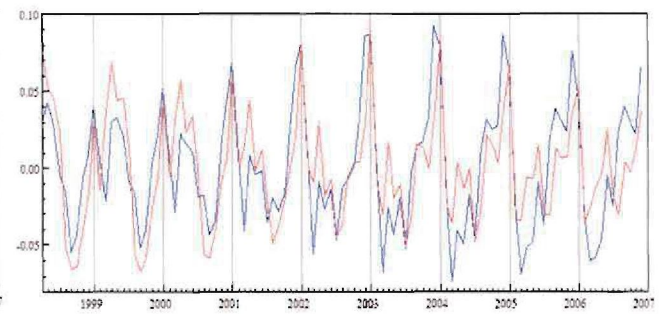
GRAPH 6.44 — WM 2 away —YM 1 away



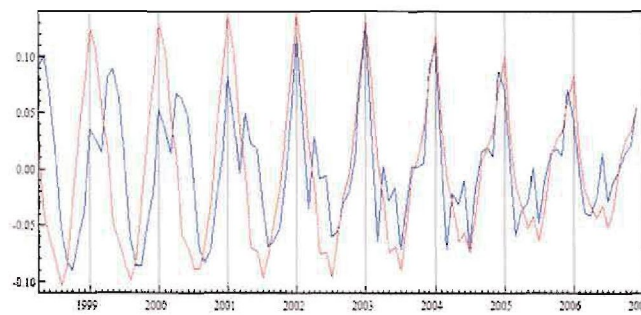
GRAPH 6.45 —WM 2 away —YM 2 away



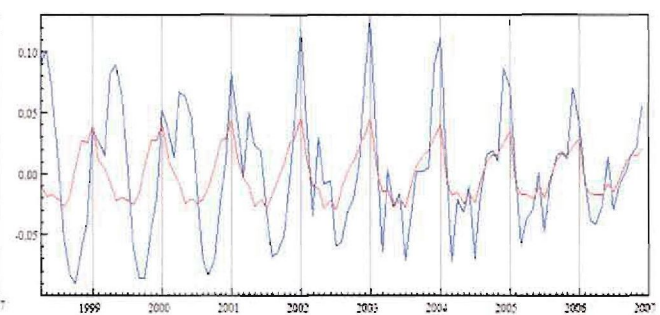
GRAPH 6.46 — WM 2 away —YM 3 away



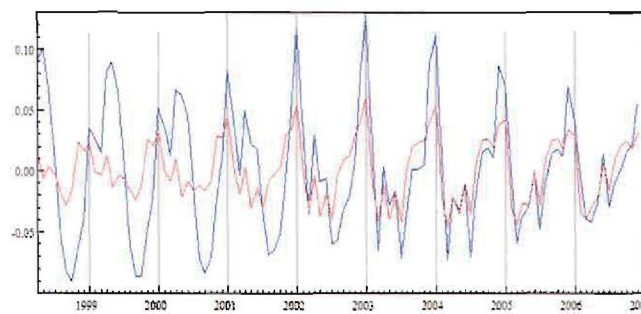
GRAPH 6.47 —WM 3 away —YMSPOT



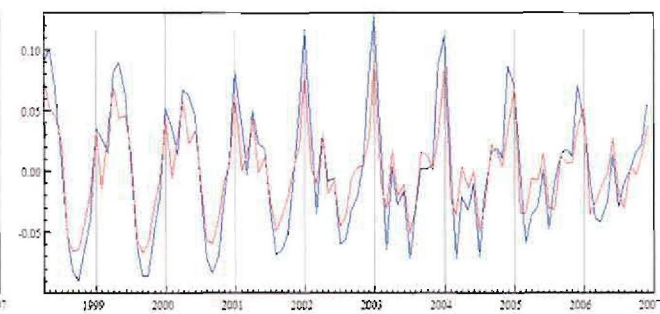
GRAPH 6.48 —WM 3 away —YM 1 away



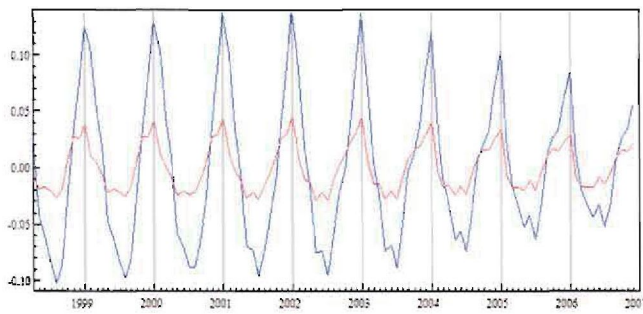
GRAPH 6.49 —WM 3 away —YM 2 away



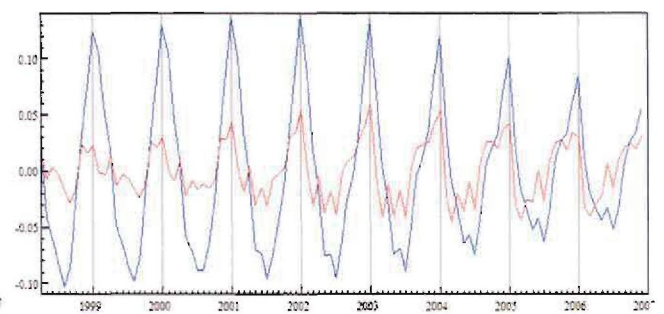
GRAPH 6.50 — WM 3 away —YM 3 away



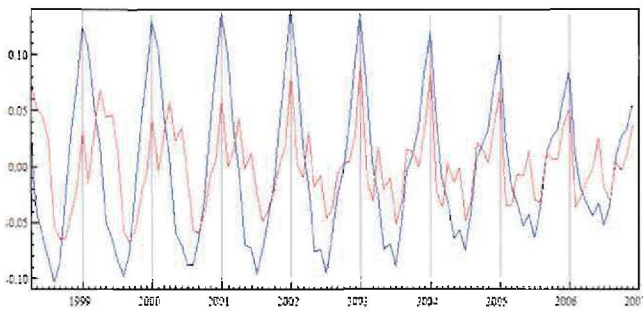
GRAPH 6.51 —YMSPOT —YM 1 away



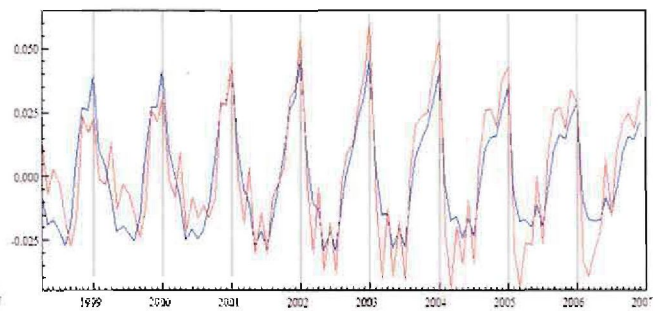
GRAPH 6.52 —YMSPOT —YM 2 away



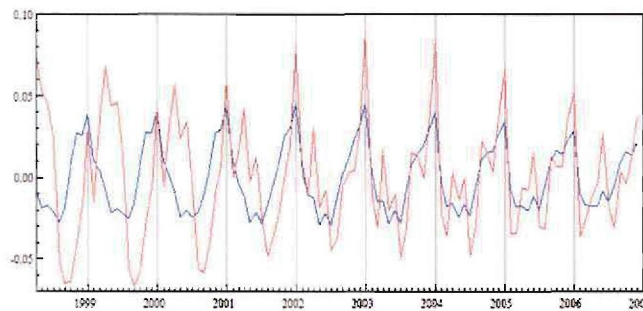
GRAPH 6.53 —YMSPOT —YM 3 away



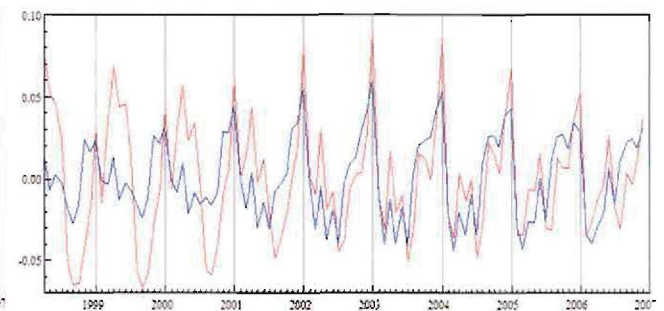
GRAPH 6.54 —YM 1 away —YM 2 away



GRAPH 6.55 —YM 1 away —YM 3 away



GRAPH 6.56 —YM 2 away —YM 3 away



Upon examination of the seasonal factors series it can be seen that the yearly seasonal patterns are much more volatile than the long-run trends. Therefore, it is not surprising that the seasonal factors series of the preliminary pairs are generally less correlated than the Henderson-curves for the same pairs.

The preliminary pairs portfolio was thus comprised of only seventeen of the original twenty-eight pairs selected initially. Table 6.15 reports the preliminary pairs portfolio selected by the various statistical and the deterministic seasonal tools.

TABLE 6.15 PRELIMINARY PAIRS PORTFOLIO

WMSPOT & WM 1 away	WMSPOT & WM 2 away	WMSPOT & WM 3 away	WMSPOT & YMSPOT
WMSPOT & YM 1 away	WM 1 away & WM 2 away	WM 1 away & YMSPOT	WM 1 away & YM 1 away
WM 1 away & YM 2 away	WM 2 away & WM 3 away	WM 2 away & YM 1 away	WM 2 away & YM 2 away
WM 2 away & YM 3 away	WM 3 away & YM 3 away	YMSPOT & YM 1 away	YMSPOT & YM 2 away
YM 1 away & YM 2 away			

The final pairs portfolio was determined by repeating the same validation process for UCM on the preliminary pairs selected using seasonal factors as a pair selection tool. After assembling the final pairs portfolio, the pairs trader will need to ensure that all of the selected pairs are robust in their co-moving behaviour.

### 6.3.3 Conclusion

The results of the Henderson-curve data are consistent with those of the statistical tests. Accordingly, all twenty-eight preliminary pairs were again selected as long-run pairs. The reason for these results is that all of the maize series have very similar underlying long-run trends. The seasonal factors series results on the other hand indicated that only seventeen of the original twenty-eight preliminary pairs should be regarded as long-run pairs. The correlation between these series is on average lower than that of the Henderson-curves.

Making use of the Henderson-curve and seasonal factors series as pairs selection tools ensured that not only were the statistical properties of the data taken into consideration, but also that pairs were selected based on their seasonal history. The next section discusses the final process in selecting pairs by making use of UCM.

## 6.4 Pairs Selection: Stochastic Seasonality Measures

### 6.4.1 Introduction

UCM can be employed in two ways: to indicate the hidden seasonal patterns in the maize data, and to identify pairs. Section 6.4.2 explains how UCM was employed to indicate hidden seasonal patterns as a pairs selection tool.

## 6.4.2 Testing for pairs with UCM: Methodology

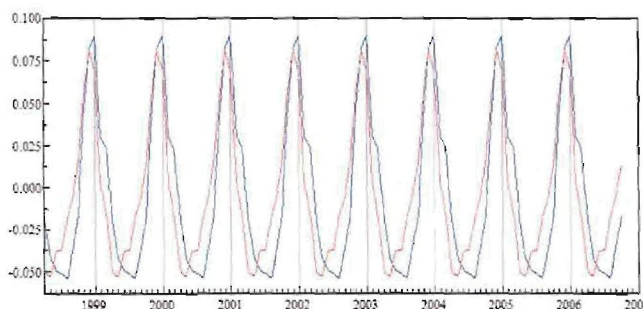
As was the case with the deterministic seasonal patterns, the correlation coefficients between the hidden seasonal series were used as an indication of the co-movement between the different maize series. If the correlation coefficient between two maize series was less than 0.7 it was no longer regarded as a valid pair. Of the seventeen remaining pairs, the correlation test indicated that only twelve have correlation coefficients of 0.7 and more. This means that five of the preliminary pairs selected by the previous seasonality tests were not included in the final portfolio.

The pair with the strongest statistical relationship is WM 1 away & YM 1 away (0.983), followed by WM 2 away & YM 2 away (0.975), and then WMSPOT & YMSPOT. It is important to remember that this is the correlation coefficient between the two maize series' hidden seasonal series not between the actual maize series. Conversely, the pairs with the weakest relationships are YMSPOT & YM 2 away (0.404), followed by WM 2 away & YM 3 away (0.448), and then WM 2 away & WM 3 away (0.526). Table 6.16 reports the cross-correlation matrix for the different hidden seasonal series,<sup>83</sup> and graphs 6.57 to 6.75 illustrate the comparisons of the final pairs' UCM series.<sup>84</sup>

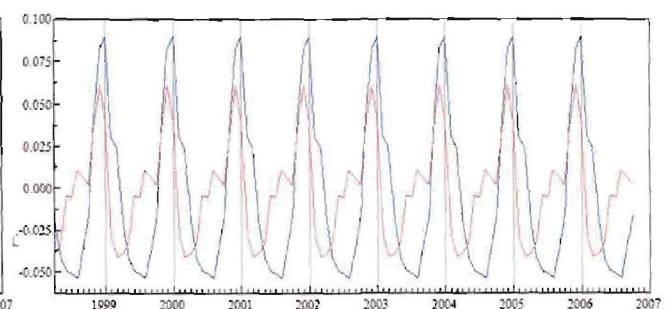
TABLE 6.16 CORRELATION COEFFICIENTS BETWEEN SEASONAL FACTORS SERIES

Agricultural Commodity Pair	WM SPOT	WM 1 away	WM 2 away	WM 3 away	YM SPOT	YM 1 away	YM 2 away	YM 3 away
WMSPOT	1.000	0.858	0.534	0.632	0.971	0.789		
WM 1 away	0.858	1.000	0.853		0.861	0.983	0.797	
WM 2 away	0.534	0.853	1.000	0.526		0.842	0.975	0.448
WM 3 away	0.632		0.526	1.000				0.954
YMSPOT	0.971	0.861			1.000	0.823	0.404	
YM 1 away	0.789	0.983	0.842		0.823	1.000	0.824	
YM 2 away		0.797	0.975		0.404	0.824	1.000	
YM 3 away			0.448	0.954				1.000

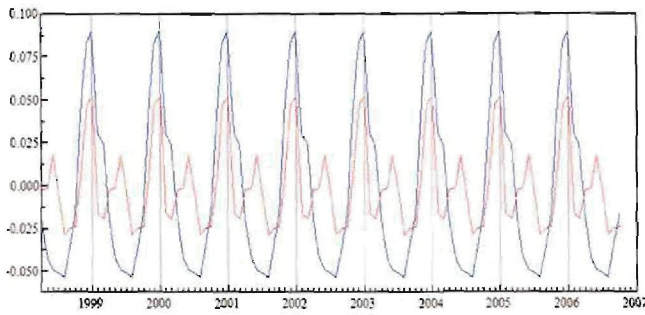
GRAPH 6.57 —WMSPOT —WM 1 away



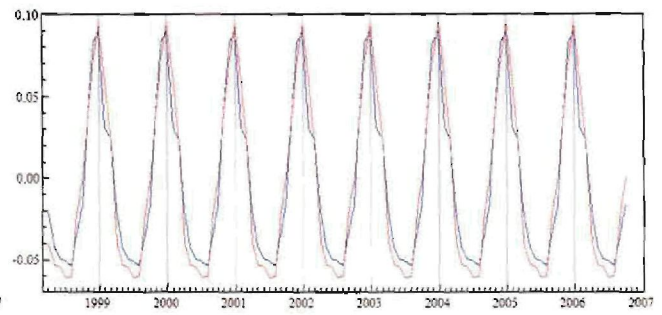
GRAPH 6.58 —WMSPOT —WM 2 away



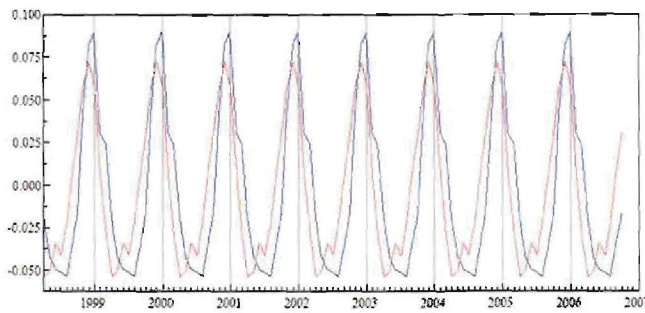
GRAPH 6.59  $\rightarrow$ WMSPOT  $\rightarrow$ WM 3 away



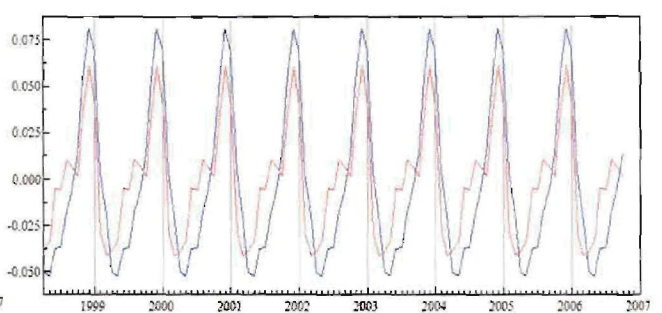
GRAPH 6.60  $\rightarrow$ WMSPOT  $\rightarrow$ YMSPOT



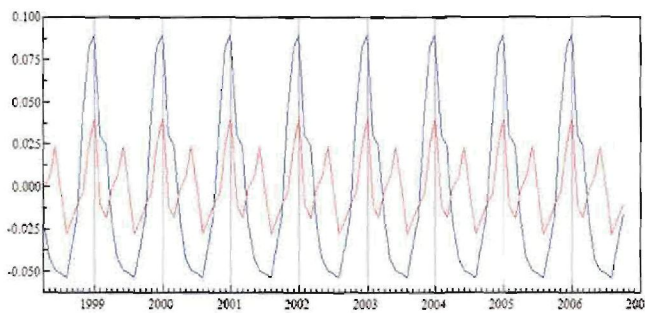
GRAPH 6.61  $\rightarrow$ WMSPOT  $\rightarrow$ YM 1 away



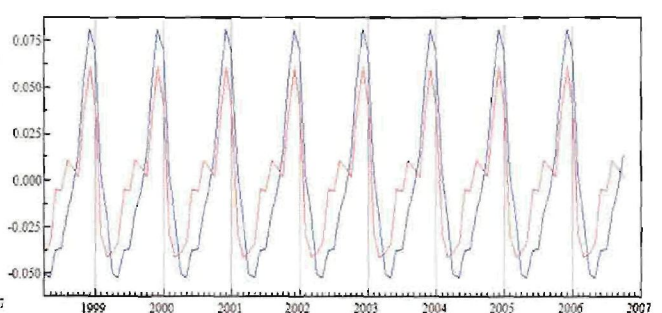
GRAPH 6.62  $\rightarrow$ WM 1 away  $\rightarrow$ YM 2 away



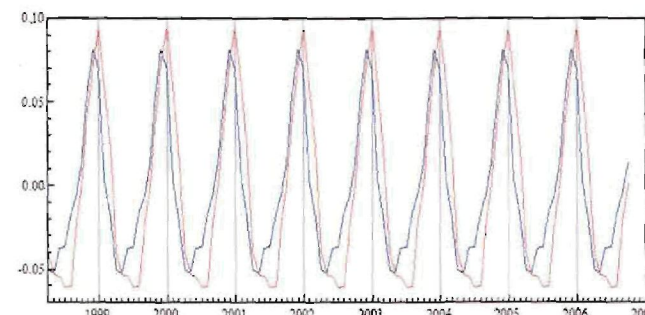
GRAPH 6.63  $\rightarrow$ WMSPOT  $\rightarrow$ YM 3 away



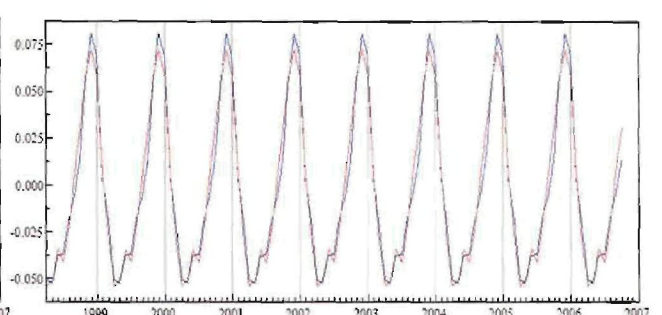
GRAPH 6.64  $\rightarrow$ WM 1 away  $\rightarrow$ WM 2 away



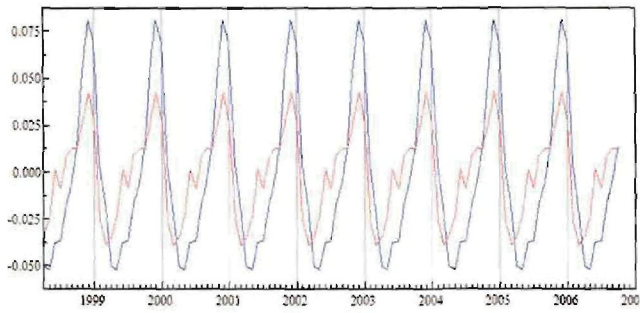
GRAPH 6.65  $\rightarrow$ WM 1 away  $\rightarrow$ YMSPOT



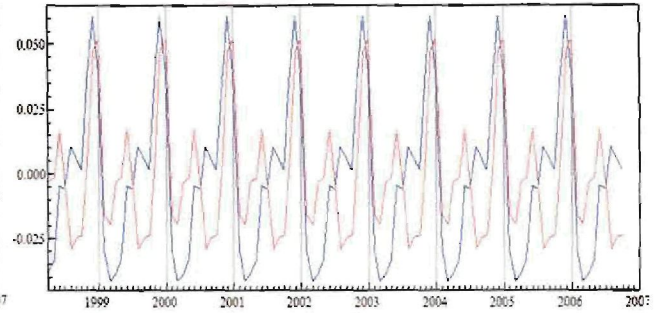
GRAPH 6.66  $\rightarrow$ WM 1 away  $\rightarrow$ YM 1 away



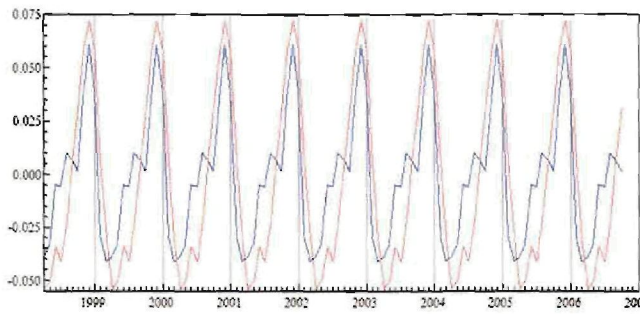
GRAPH 6.67 →WM 1 away →YM 2 away



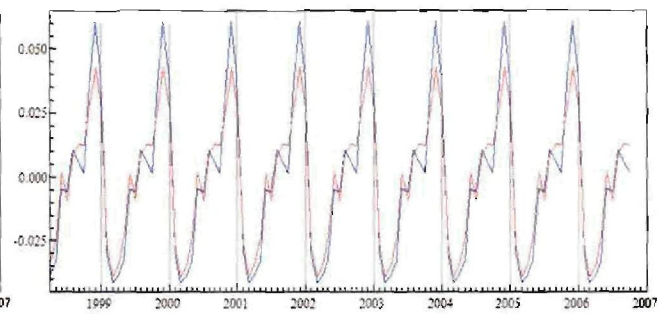
GRAPH 6.68 →WM 2 away →WM 3 away



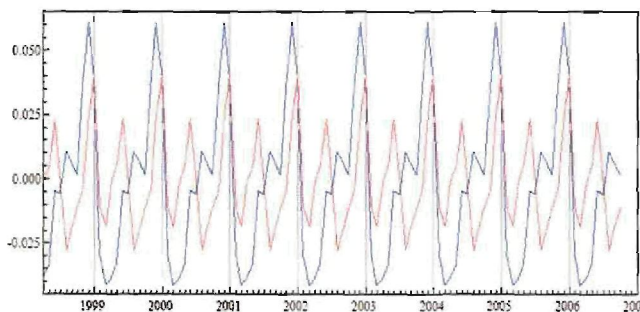
GRAPH 6.69 →WM 2 away →YM 1 away



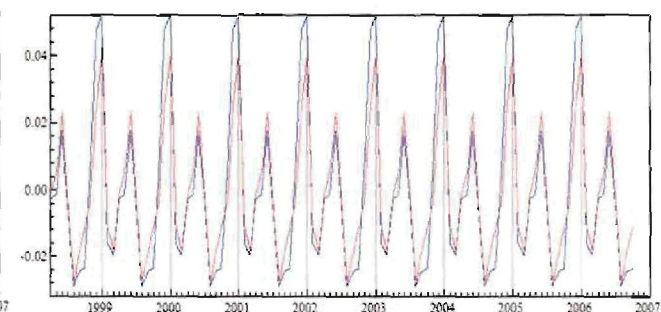
GRAPH 6.70 →WM 2 away →YM 2 away



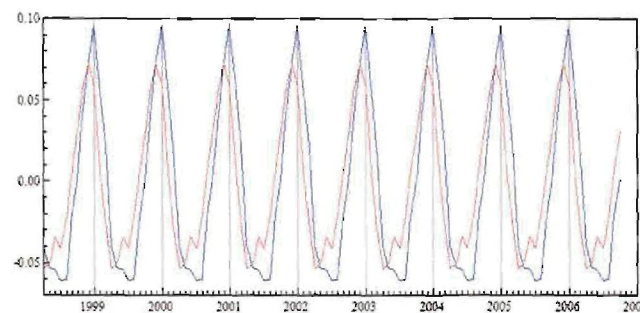
GRAPH 6.71 →WM 2 away →YM 3 away



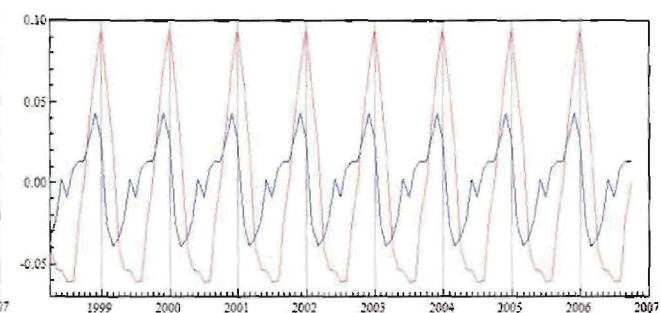
GRAPH 6.72 →WM 3 away →YM 3 away



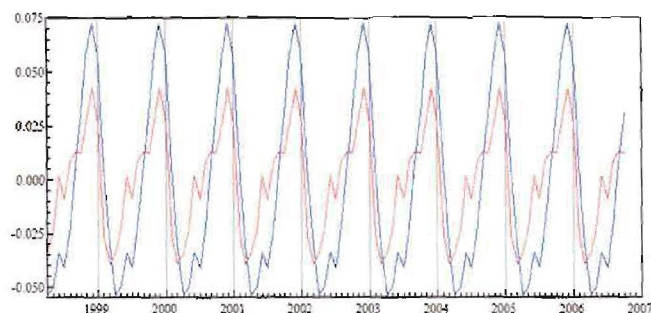
GRAPH 6.73 →YMSPOT →YM 1 away



GRAPH 6.74 →YMSPOT →YM 2 away



GRAPH 6.75 —YM 1 away —YM 2 away



The final pairs portfolio thus was comprised of only twelve of the original seventeen pairs selected during the previous seasonality tests. Table 6.17 reports the final pairs portfolio selected during the final phase of the pairs selection process.

TABLE 6.17 FINAL PAIRS PORTFOLIO

WMSPOT & WM 1 away	WMSPOT & YMSPOT	WMSPOT & YM 1 away	WM 1 away & WM 2 away
WM 1 away & YMSPOT	WM 1 away & YM 1 away	WM 1 away & YM 2 away	WM 2 away & YM 1 away
WM 2 away & YM 2 away	WM 3 away & YM 3 away	YMSPOT & YM 1 away	YM 1 away & YM 2 away

The final pairs portfolio was test traded (the results will be reported in chapter 7) by making use of fundamental analysis as well as technical and seasonal analysis. Since seasonal patterns change over time, it is imperative that these tests are performed regularly in order to ensure that all of the selected pairs are still valid pairs.<sup>85</sup>

### 6.4.3 Conclusion

The tests for hidden seasonal patterns in the data revealed that these hidden seasonal patterns give an accurate description of the actual maize data. Given this result, UCM was used effectively as pairs selection tool here. Since the use of UCM as pairs selection tool is the final step in the pairs selection process, the next step will be for the trader to trade the final selection of pairs. After performing this rigorous selection process, the trader can be sure that all the pairs included in the final selection will hold as long-run pairs and continue to revert back to a common mean in future.

<sup>85</sup> To ensure that all of the current pairs in a portfolio are still valid pairs, all of the statistical and seasonal tests reported in this thesis should be repeated for each pair on a monthly basis.

## 6.5 Summary

The pairs selection process is critical to the success of the commodities pairs trader. However, it is not enough only to establish whether two commodities have a statistical relationship over the immediate short-run. Rather, the validity of pairs need be tested over the long-run, in the search for maize series that form stable enough relationships to be included in a pairs portfolio. The first step in the pairs selection process should therefore be to test for co-integrating relationships between all the combinations of white and yellow maize over the various time periods. To ensure that the chosen pairs possess a strong enough statistical relationship, only those combinations with co-integrating relationships are to be considered long-run pairs.

Testing for long-run pairs by means of co-integration resulted in the assembling of twenty-eight pairs comprised of white and yellow maize futures contracts over various periods.<sup>86</sup> After establishing which pairs were viable long-run pairs, the validity of the selected pairs was determined by subjecting each pair to Pearson correlation tests and Spearman rank correlation tests. These tests were performed in order to cross-check the co-integration results.

Tests regarding Pearson correlation were of two sets. The first set of tests made use of the natural logarithmic returns in the daily closing prices of the series tested. These tests indicated that most of the commodity combinations (long-run pairs) have a strong statistical relationship between them. The second set of tests established how strong the statistical relationship between the residuals of the chosen pair is, which is more critical than establishing strength only. The results of the Pearson correlation tests performed with the residuals data indicated that the twenty-eight pairs (selected during the co-integration tests) were to be included in the pairs portfolio.

Since the results of the Pearson correlation tests confirmed the stability of the statistical relationship between the commodities constituting the long-run pairs, the selected pairs were further verified by means of the Spearman rank correlation test. These tests were conducted in the same manner as the tests for Pearson correlation.

---

<sup>86</sup> Although products such as wheat, sunflower and soybeans are also traded on SAFEX, these products do not have the necessary liquidity to make such a pair a viable option.

The Spearman rank correlation results corroborate the findings of the co-integration and Pearson correlation tests. As before, the first set of tests made use of the natural logarithmic returns in the daily closing prices of all the commodity series. These tests confirm that most of the commodity combinations (long-run pairs) have a strong statistical relationship between them. The second set of the Spearman rank correlation tests was performed with the residuals data. These results support the Pearson correlation residuals results, indicating that the twenty-eight pairs selected during the co-integration tests were to be included in the pairs portfolio.

All the tests performed up to this point indicated that the statistical relationship between the commodity combinations warranted their inclusion in a pairs portfolio. Testing for the co-movement between these series based on their seasonal patterns was the final step in establishing the nature of the relationship between the commodities that constitute each pair.

Although Beta does not give a clear indication of the strength of the statistical relationship between the commodities that makes a pair, it too was an indication of the co-movement of the two commodities. As before, two sets of data were used in testing for the relationship between the commodities forming each pair. The results of the first set of tests (using the logarithmic returns data) show that the dependent variables react similarly to their benchmark commodities, given exposure to the same set of external events. The results regarding the residuals data confirms the direction of the co-movement between the dependent commodities and their benchmarks, but indicates greater movement on the part of the dependent commodity (given exposure to the same set of external events).

After taking all the above results into consideration, the preliminary pairs portfolio was subjected to tests for deterministic seasonality. Two tests were performed to this effect. First, the correlation coefficients between the various maize series' Henderson-curves were considered. The results confirmed the results of the statistical tests. Second, after measuring the co-movement between the pairs' Henderson-curves, the co-movement between their seasonal factors series was evaluated in the same manner. Some of these

results contradict the previous results. According to the seasonal factors results, only seventeen of the original twenty-eight pairs reached the final portfolio.

After testing for pairs by means of deterministic seasonality, tests were performed to validate the pairs further by means of stochastic seasonal models. Once again, the correlation coefficients between the various maize series' hidden seasonal series were used as an indication of the co-movement between them. The results of these tests reveal that the final portfolio was to include only twelve of the original twenty-eight pairs. This final pairs portfolio was test traded by making use of historical data. Chapter 7 will report on the results of the test trades that were performed by making use of the tools that have been discussed in the preceding chapters. These include fundamental analysis (chapter 2), technical analysis (chapter 3) and seasonal pattern analysis (chapters 4, 5 and 6).

## CHAPTER 7

# Managing an Agricultural Commodities Portfolio with Pairs Trading

*The better your dealing skills, the more chance you have of making profits.*

(McHattie, 1996:134)

### 7.1 Introduction

Agricultural commodity traders are active in the market for several reasons. While some traders are only interested in hedging their positions, others follow a speculative approach in the quest for realising sizeable profits. Whatever the motive to trade, all traders will agree that, although there are numerous trading strategies available to them, they are always on the lookout for better and more effective strategies (Van Zyl, 2006).

One such strategy to be explore further is pairs trading. As previously mentioned, this thesis explores pairs trading as an option for agricultural commodities traders in South Africa. The trading strategy is mainly speculative in nature, and will best suit those traders who are interested in realising a profit from managing an agricultural commodities portfolio (Van Zyl, 2006). Although pairs trading can be regarded as a speculative trading strategy, it is possible to manage such a portfolio as a hedging tool (Whistler, 2004:109). In chapter 6, the relationships between the various commodities series were analysed in order to assemble an agricultural commodities pairs portfolio. This chapter investigates the viability of using pairs trading as a profitable option for the agricultural commodities trader. In order to maximise profit, the pairs trader will have to utilise the information covered in chapters 2 through 6.<sup>87</sup>

Section 7.2 below commences with a list of the pairs in the final portfolio that were traded, while section 7.3 covers the methodology followed in managing an agricultural

---

<sup>87</sup> It is important to note that any combination of analysis tools can be used, and that this is based on the individual trader's preferences. The tools used in this thesis can be replaced with similar tools without affecting the profitability of the portfolio on condition that the statistical relation between the maize series comprising each pair holds. This is especially true in the case of the technical tools used for the test trading process in this thesis.

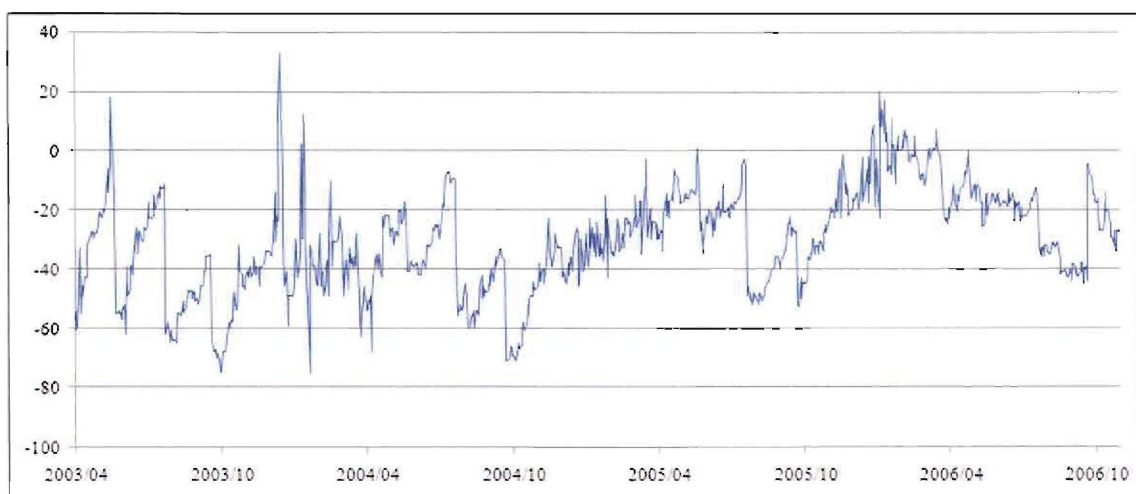
commodities pairs portfolio. This section also elaborates on the assumptions mentioned in chapter 1, regarding the management of an agricultural commodities portfolio by means of a pairs trading strategy. This is done to facilitate the agricultural commodities pairs trader in replicating the steps followed during the test trading process reported in this thesis. Thereafter, Section 7.4 reports on the results of the test trades.

## 7.2 Data

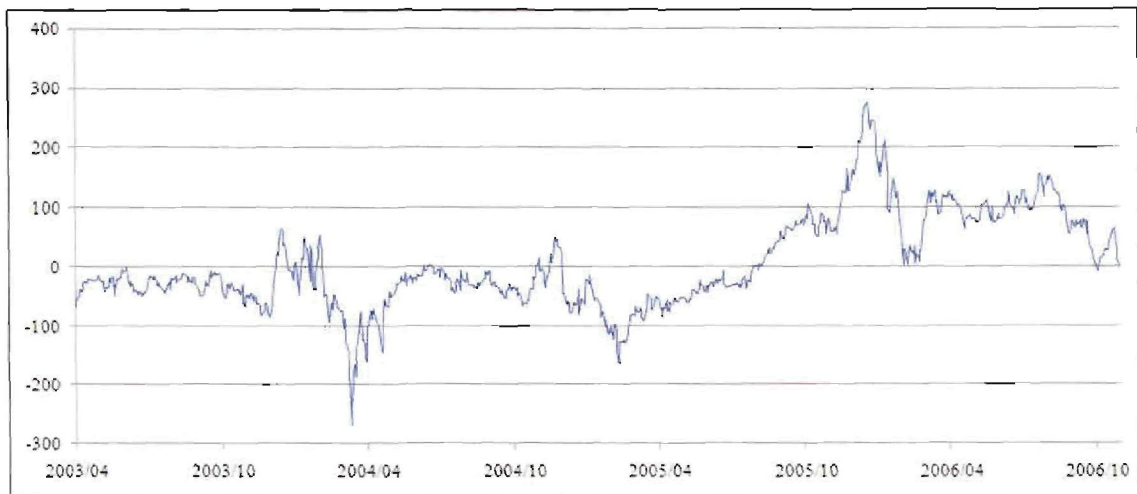
The final pairs portfolio consists of the following pairs: WMSPOT & WM 1 away, WMSPOT & YMSPOT, WMSPOT & YM 1 away, WM 1 away & WM 2 away, WM 1 away & YMSPOT, WM 1 away & YM 1 away, WM 1 away & YM 2 away, WM 2 away & YM 1 away, WM 2 away & YM 2 away, WM 3 away & YM 3 away, YMSPOT & YM 1 away, and YM 1 away & YM 2 away.

The test period commenced on the 26th of March 1998, and ended on the 31st of October 2006. Graphs 7.1 to 7.12 give a graphical illustration of the data series used during the test trading process. In order to illustrate the volatility of the pairs traded during the test trading process these graphs only depict the difference between the two maize series comprising each pair. Graph 7.1, for example, shows the difference between the two series WMSPOT and WM 1 away prices expressed as a rand per metric tonne value.

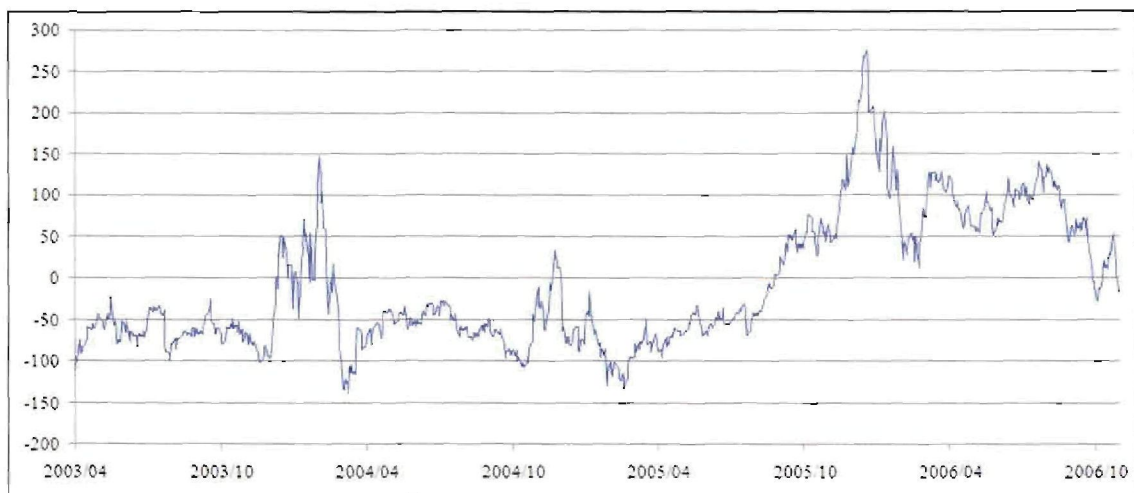
GRAPH 7.1 WMSPOT – WM 1 away



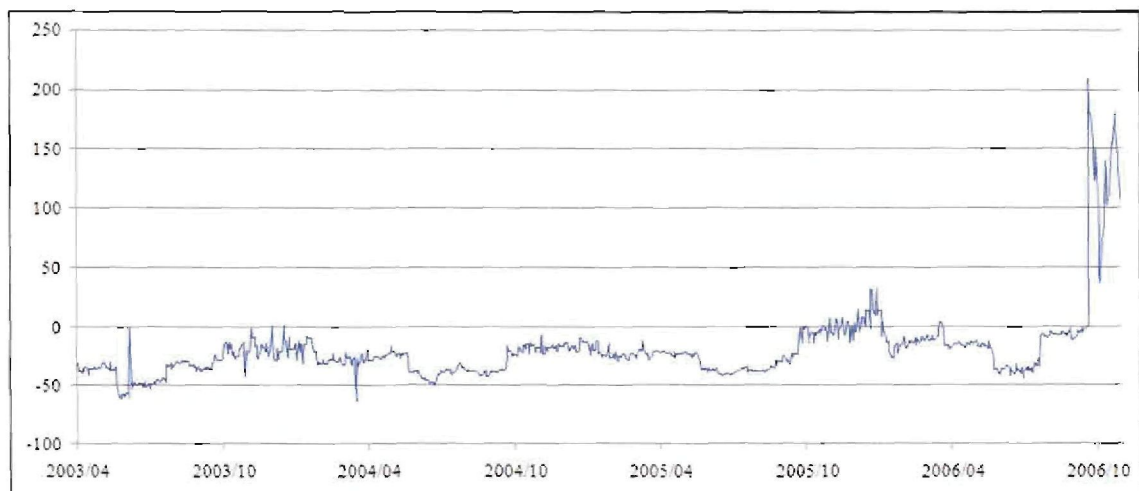
GRAPH 7.2 WMSPOT – YMSPOT



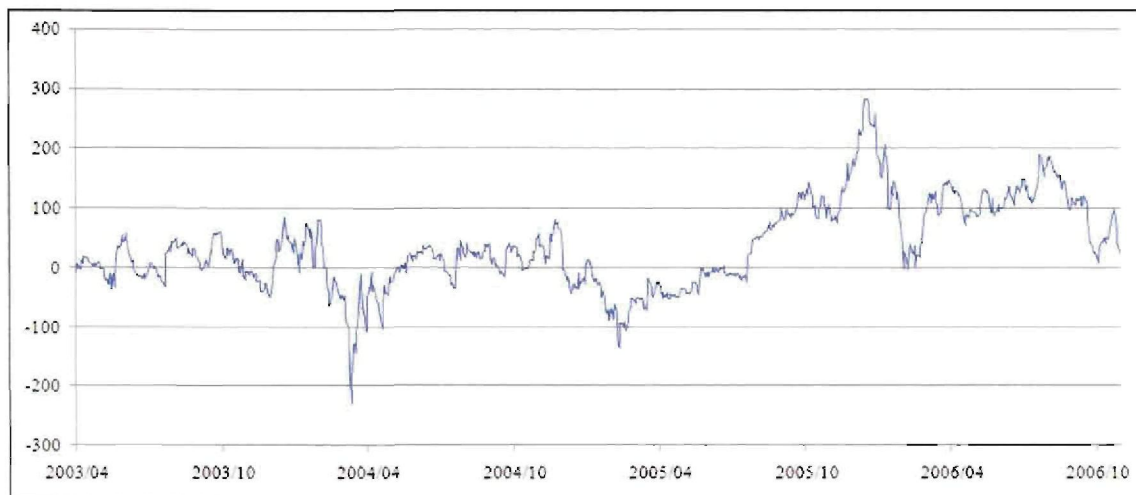
GRAPH 7.3 WMSPOT – YM 1 away



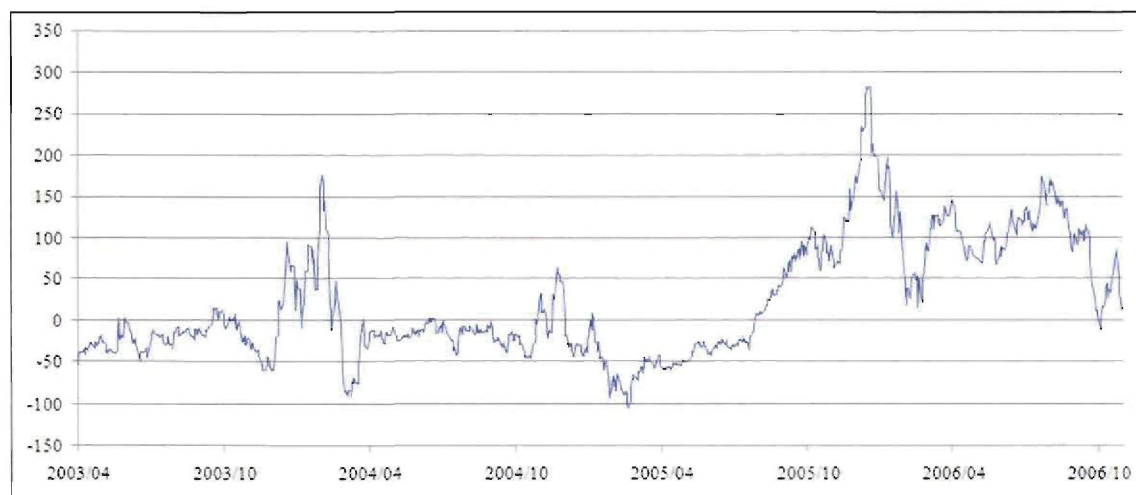
GRAPH 7.4 WM 1 away – WM 2 away



GRAPH 7.5 WM 1 away – YMSPOT



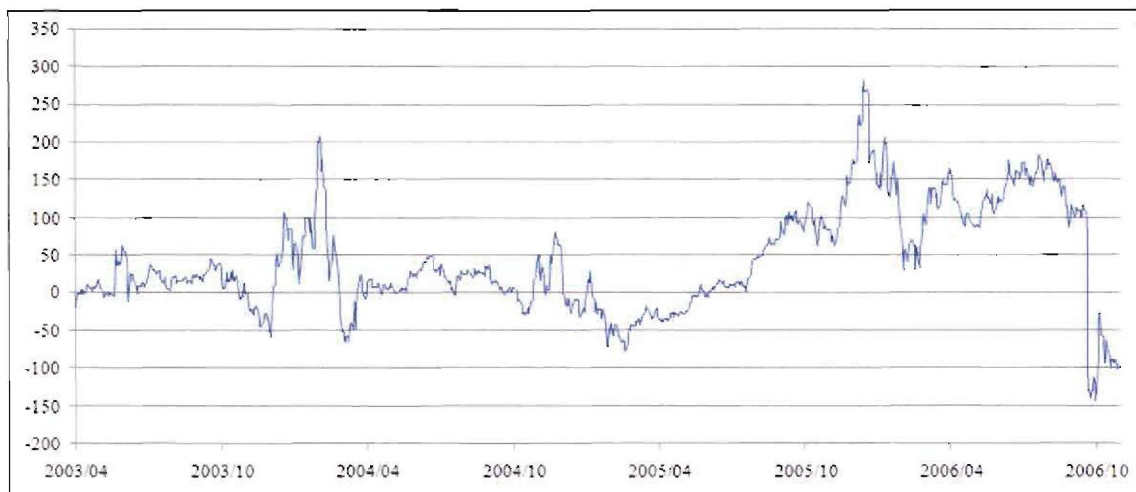
GRAPH 7.6 WM 1 away – YM 1 away



GRAPH 7.7 WM 1 away – YM 2 away



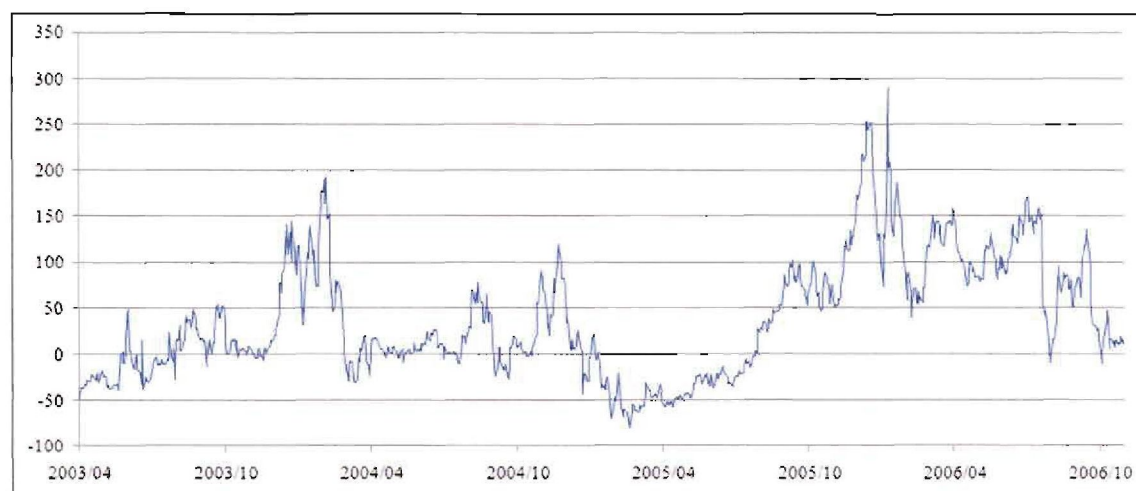
GRAPH 7.8 WM 2 away – YM 1 away



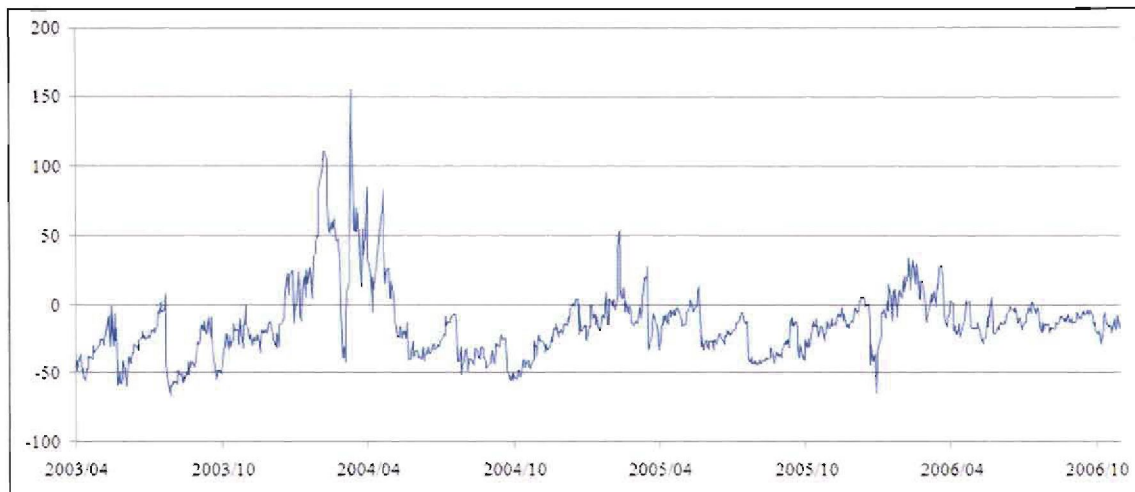
GRAPH 7.9 WM 2 away – YM 2 away



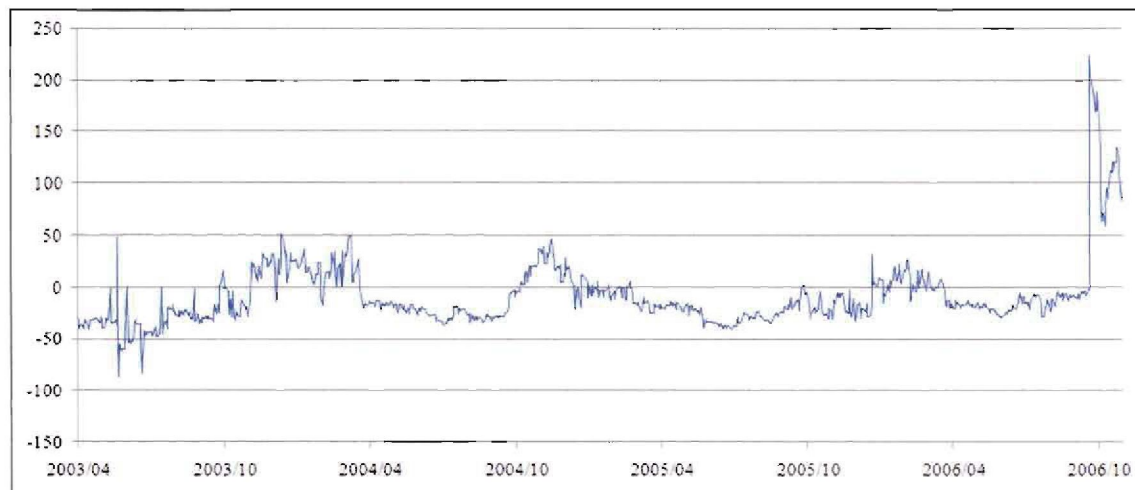
GRAPH 7.10 WM 3 away – YM 3 away



GRAPH 7.11 YMSPOT – YM 1 away



GRAPH 7.12 YM 1 away – YM 2 away



In order to assess the profitability of the final pairs portfolio, each pair was test traded by using fundamental analysis, technical analysis, and the seasonal patterns information.<sup>88</sup> Section 7.3 reports on the steps followed during the test trading process.

## 7.3 Managing a Pairs Portfolio

### 7.3.1 Introduction

When managing a pairs portfolio it is important to take cognisance of all the factors that play a role in determining the prices of the products to be traded. There is more than one

<sup>88</sup> Please refer to chapters 2 to 6 for information regarding these trading tools.

way to determine what these prices will do in the future. When the trader considers fundamental factors in determining future prices, he focuses on the economic forces of supply and demand that affect agricultural commodity prices. In doing so, the trader will uncover the intrinsic value of the particular agricultural commodity. In knowing what the intrinsic value of the agricultural commodity is, the trader will be able to determine whether the agricultural commodity is overpriced or under priced.

When the trader makes use of technical analysis in determining future prices, he will concentrate on the study of market movements only. Both technical and fundamental analysis attempts to solve the same problem; determining the direction of future price movements. For the purpose of this project, both technical and fundamental analysis are complemented by making use of seasonal patterns in the data as a third check regarding the determination of future price movements. Section 7.3.2 discusses the assumptions made regarding the execution of each trade, as well as the methodology followed during the test trading process.

For the purpose of this project, two approaches were followed. The first consists of test trading the final pairs portfolio by taking fundamental analysis and technical analysis into consideration in arriving at the entry and exit points. This approach allows for all these indicators to be used together. In the second approach, five of the technical analysis tools discussed in chapter 3, namely, the RSI, the Stochastic Oscillator, MA, Bollinger bands, and MACD were employed in isolation in determining entry and exit points.<sup>89</sup>

Although fundamental factors play an important role in the determination of maize prices, much more emphasis was placed on the signals derived from the individual technical analysis tools. This approach was followed because of the difficulty of simulating the impact that the fundamental factors have had on the price level of the maize series long after these events occurred. However, it is to be noted that more emphasis would be placed on fundamental analysis when replicating this process in practice (real-time), potentially yielding even better results.

---

<sup>89</sup> Please note that CMO was not available on the version of Market Master used for identifying entry and exit points.

Finally, note that the results of all the trades are available on the accompanying CD. In order to simplify the final results for both approaches only the trades that were executed are reported on the CD. Also, no reference was made to the specific indicators that affected the trades on the date of execution. This reporting style was applied for both testing approaches. Please refer to the readme document for information regarding these results.

## 7.3.2 Methodology

### 7.3.2.1 Introduction

Before the pairs trader can commence with the management of an agricultural commodities portfolio by implementing a pairs trading strategy, it is imperative that he is familiar with the assumptions and trading rules to be followed to this end. Once familiar with the assumptions regarding the trading process, it is also important to know how the pairs will be traded.

Therefore, section 7.3.2.2 briefly discusses all assumptions regarding the trading process. Thereafter, section 7.3.2.3 describes the trading process, while the results of this process are reported in section 7.4.

### 7.3.2.2 Assumptions and Trading Rules

#### 7.3.2.2.1 Introduction

Three assumptions are made regarding the trading process. The first assumption clarifies all information pertaining to entry and exit points for a trade. The second assumption addresses the identification of buy and sell signals for each of the tools used. The third assumption clarifies the cost structure of all the trades made during the test trading period.

#### 7.3.2.2.2 Entry and Exit Points

With pairs trading, each entry and exit point is comprised of going long in (buying) one maize series, whilst simultaneously selling the other short. If the trader intends to

remain in the market permanently, the signal for exiting one position will be considered an entry point for the opposite position.

When entering the market, the trader buys maize series A, while simultaneously selling short maize series B. When exiting the trade, maize series A is sold whilst buying back maize series B. In doing this, the original position is closed. However, when exiting the trade, the original unit of maize series A is sold whilst short selling an additional unit of maize series A. Similarly, when buying back the original unit of maize series B, an additional unit of maize series B is bought.

By doing this, the trader remains active in the market permanently, making the maximum amount of profit from all increases and decreases in the maize series' prices. Alternatively, the trader can enter the market at a buy signal, exit at the first sell signal, and only re-enter the market at the next buy signal. For the purpose of this project, the trader remains in the market on a permanent basis. It is important to note that the trader is also required to perform additional transactions in order to remain in the particular positions. By way of explanation, the following example is given: the trader enters into a pairs trade on the 1st of September 2003. On this date, he buys ten WM 1 away contracts whilst short selling ten WM 2 away contracts. By the 21st of September, the current WMSPOT series expires, causing the WM 1 away contract to become the new WMSPOT contract, and WM 2 away contract to become the new WM 1 away contract.

By the 21st of September 2003, no buy or sell signals are registered by any of the indicators. For the trader to retain his current position after the 21st of September 2003 he is forced to sell his December 2003 white maize contract (old WM 1 away contract) and buy a March 2004 white maize contract (new WM 1 away contract). Similarly, he has to buy back his March 2004 white maize contract (old WM 2 away contract) whilst short selling another May 2004 white maize contract (new WM 2 away contract).

It is important to note that the data series used were not compiled by making use of constant contract months. The reason for this is that positions taken in these months might be exercised by counter parties at any time.

### 7.3.2.2.3 Identifying Entry and Exit Points

In order to test for entry and exit points, the price difference between the two maize series constituting the pair is used. Entry and exit points are determined by making use of fundamental analysis, the RSI, the Stochastic Oscillator, MA, Bollinger bands, and MACD.

When regarding fundamental analysis, entry and exit points are determined by the factors that influence the supply and demand of the underlying commodity. Since the same event will affect each individual maize series in a pair differently at different times of the year, entry and exit signals are considered at the time when the intended deal is to be placed.<sup>90</sup>

In utilising the RSI as tool to identify entry and exit points the following assumptions are made. An entry point is registered once the RSI indicator crosses the bottom limit (30 percent line) from above, while an exit point is registered when the RSI indicator crosses the top line limit (70 percent line) from below. An entry point means the maize series A should be bought, while short selling maize series B. An exit point indicates that maize series A should be sold, while maize series B is bought back. For the purpose of this project, a five-day RSI is used in testing for entry and exit points.<sup>91</sup>

In making use of the Stochastic oscillator, entry and exit points are registered as follows. Entry signals are registered when the indicator is between the 10 and 15 levels, and the %K line crosses the %D line from below. Similarly, exit signals are registered when the indicators is between the 85 and 90 levels, and the %K line crosses the %D line from above.<sup>92</sup>

Even though moving averages can be utilised in many different ways to indicate entry and exit points, moving averages are used in conjunction with the original price series

---

<sup>90</sup> Please refer to chapter 2 for a discussion on the factors that play a role in the determination of the price of agricultural commodities.

<sup>91</sup> Please refer to section 3.2.4 for a discussion on RSI. Although other limits and time frames could have been used, preliminary tests revealed that the five-day RSI with 30 percent and 70 percent limits renders the best results for identifying oversold and overbought conditions. Please refer to section 7.2.3.4 for an explanation of the preliminary tests' results.

<sup>92</sup> Please refer to section 3.2.5 for a discussion on the Stochastic oscillator.

for the purpose of this thesis. Entry points are identified when the original price crosses the moving average from below, while exit points are identified when the original price series cross the moving average from above. A moving average of 5 days was used in determining entry and exit points.

In using Bollinger bands, an entry point is registered once the price crosses the bottom band line from below after breaking out of the two bands' range at the bottom. Conversely, an exit point is identified when the price crosses the top band line from above after breaking out of the band range at the top.<sup>93</sup>

When using MACD, an entry point is generated when the MACD line rises above the signal line from below. In contrast, an exit point is generated when the MACD line crosses the signal line from above.<sup>94</sup>

Note that since end-of-day closing prices are used, the signal only registers at the end of the trading day. This is true for all four of the technical indicators used to identify entry and exit points. The position, therefore, can only be taken in the morning of the following day. For the purpose of this thesis, these transactions are recorded at the closing price of the day, following the day in which the signal was registered.

#### 7.3.2.2.4 Brokerage and Margins

When trading agricultural commodity pairs, the trader is subjected to three different costs. First, for every deal placed by the trader, margin payments are made to cover any losses incurred during the trading process. Second, the trader has to pay brokerage on all the deals placed. Third, an opportunity cost is incurred when the trader's money is in the margin account, where it earns a market related interest rate. It is possible that, if the same amount was invested elsewhere, more interest could have been earned for the same risk exposure.

The following assumption is made regarding the first cost. It is assumed that an initial margin of R 10 000 is needed to buy or sell one futures contract. To enter into a pairs

---

<sup>93</sup> Please refer to section 3.3.3 for a discussion on Bollinger bands.

<sup>94</sup> Please refer to section 3.3.4 for a discussion on the MACD indicator.

trade (if one futures contract is bought, while another is sold short) where maize series A is a white maize series and maize series B a yellow maize series, an initial margin of R 20 000 needs to be available in the trader's SAFEX account. When entering into a pairs trade, where both maize series are either white or yellow, the trader only needs to hold R 4 000 in his SAFEX account (Malan, 2007).

Since the margin (amount of money) available to each trader differs according to his individual mandate, it is also assumed that only one contract is bought and one contract sold short in taking a position. This is done to simplify all the reports on the profitability of each pair. Furthermore, it is important to note that the size of the initial margin will vary daily, as the prices of the futures contracts change. The pairs trader will thus receive a margin call once the value of his portfolio falls below a certain amount.

Regarding the second cost, it is assumed that the agricultural commodities trader is a registered broking firm, and that futures contracts can be bought and sold at 75 cents per tonne. If the agricultural commodities trader is not a registered broking firm, but a private market participant, broking fees of up to R1,50 per tonne can be paid for buying or selling a futures contract on SAFEX (Van Zyl, 2006). Brokerage costs are paid in two instances: first, when the trader buys and sells futures contracts, based on the signals indicated by the various analysis tools, and second, when the trader buys and sells futures contracts in order to remain in the same series.

The third cost, opportunity cost, cannot be determined since the trader might have invested the money used for trading anywhere else. It is, however, important to take cognisance of this third cost, since the returns on this portfolio can be compared to rates of return on other investments.

### 7.3.2.3 Steps Followed during the Test Trading Process

In testing the profitability of the final pairs portfolio selected in chapter 6, two different approaches were followed. The first approach entailed testing for entry and exit points, by employing all the technical analysis tools discussed in section 7.3.2.2.3 above.

In this approach, an entry point was only confirmed once two technical analysis tools confirmed a day on which the trade should be executed. Note that it is only after a second entry/exit point from a different technical analysis tool was registered, that the deal was executed. This was done to ensure that only a strong enough entry/exit signal would result in a deal being placed.

In the second approach, each of the abovementioned technical analyses tools were utilised in isolation from each other to test for entry and exit points. This means that no seasonal analyses tools are taken into consideration, and that each entry and exit point is identified by placing more weight on the technical analysis tool used at a time. It should further be noted that, once an entry signal is registered, other entry signals are ignored until an exit signal is registered, and vice versa. This process is repeated for all the technical analyses tools.

## 7.4 Results

### 7.4.1 Introduction

The results are reported in two sections. Section 7.4.2 reports on the results of the process followed using fundamental, seasonal, and technical analysis tools. Thereafter, section 7.4.3 reports on the final results achieved by only using one technical analysis tool at a time. The full trade report of all the trades that were executed for both approaches and the summary of all the results can be viewed on the accompanying CD.

### 7.4.2 Final Results of Approach One: All indicators

The final results are expressed as the rand per metric tonne value of the profit or loss made over the test period. Since one futures contract allows the holder to buy 100 metric tonnes of maize, the final profit or loss value must be multiplied by 100 in order to calculate how much money the trader would have made or lost when buying one futures contract and selling short another.

The final results of the tests performed by taking all the different indicators into consideration are given in Table 7.1. The far right-hand side column reports the amounts of money that were made or lost for each of the twelve pairs.<sup>95</sup>

TABLE 7.1 FINAL RESULTS: ALL INDICATORS

Pair	Total number of trades	Number of successful trades	Percentage of successful trades	Brokerage paid to retain positions	Profit or Loss in S.A. rand/ton
WMSPOT & WM 1 away	173	32	18.50%	-43.50	-1144.50
WMSPOT & YMSPOT	144	50	34.72%	-60.00	-55.00
WMSPOT & YM 1 away	141	49	34.75%	-43.50	-474.5
WM 1 away & WM 2 away	187	19	10.16%	-27.00	-1367.00
WM 1 away & YMSPOT	120	41	34.17%	-43.50	-810.50
WM 1 away & YM 1 away	131	48	36.64%	-27.00	-389.00
WM 1 away & YM 2 away	138	45	32.61%	-27.00	-364.00
WM 2 away & YM 1 away	139	58	41.73%	-27.00	-176.00
WM 2 away & YM 2 away	145	45	31.03%	-27.00	-460.50
WM 3 away & YM 3 away	125	52	41.60%	-27.00	-592.00
YMSPOT & YM 1 away	152	45	29.61%	-43.50	-705.50
YM 1 away & YM 2 away	152	21	13.82%	-43.50	-1481.00
Totals: All Pairs	1747	505	28.91%	-439.50	-8019.50

In making use of all the available indicators to determine entry and exit points, only 505 out of the 1747 trades made over the test period resulted in a profit. This means that only 28.91 percent of all the trades performed of this period was successful. Losses were reported for all twelve of the pairs traded over the test period.<sup>96</sup> The pair that earned the smallest loss was WMSPOT & YMSPOT (-R 55,00/tonne) followed by WM 2 away & YM 1 away (-R 176,00/tonne). The pair that realised the biggest loss for the same period was YM 1 away & YM 2 away (-R 1481,00/tonne) followed by WM 1 away & WM 2 away (-R 1367,00/tonne).

A total loss of -R 8019,50/tonne was realised over the test period. If ten contracts were traded at a time for each pair, that is, if ten contracts were bought on the one leg, while ten contracts were sold short on the other leg, a total loss of R 8 019 500,00 would have been realised.

<sup>95</sup> Please refer to the accompanying CD for a complete account of all the trades for each of the individual pairs.

<sup>96</sup> It is to be noted that the brokerage paid to retain positions were not included in the costing of each trade, but was rather added at the end of the trading period, and is reflected in the final profit or loss for each pair.

In order to realising this loss, R 1.76 million was kept in a SAFEX margin account on a permanent basis. To put this result in perspective, it is expressed as a percentage return on equity over a period of one year.<sup>97</sup> Calculated as such, R 1.76 million had to be kept in a SAFEX margin account in order to realise a loss of R 2 240 083,80 per annum.

When calculated as a rate of return on equity per annum, this strategy results in a loss of 127.28 percent on the initial capital in the SAFEX margin account. Since a pairs strategy performed with futures contracts is very risky, and a large portion of the initial margin can be lost in a very short period of time, it is more sensible to invest the money used for margin, in a risk free instrument with a higher rate of return. From the results reported in Table 7.1, it is clear that making use of all the instruments together to determine when to enter or exit a trade resulted in a large loss, and the agricultural commodities trader should not consider it as trading option.

To assess whether these pairs could be traded more successfully, each of the technical analysis tools were employed on their own. These results are discussed in section 7.4.3.

### 7.4.3 Final Results of Approach Two: Technical Analysis Tools

Note that, although fundamental analysis played a role in determining entry and exit points here, much more emphasis was rather placed on the applicable technical analysis tool's entry and exit signals.

#### 7.4.3.1 Utilising the RSI to Identify Entry and Exit Points

As was the case with the results for all the indicators used, the results for the RSI tests are expressed as the rand per metric tonne value of the profit or loss made over the test period. These results, as reported in Table 7.2, were acquired by only relying on the RSI to identify entry and exit points. As before, the far right-hand side column reports the amount of money that was made or lost for each of the twelve pairs over the test period.<sup>98</sup>

---

<sup>97</sup> To calculate the return on equity per annum the total profit is divided by 3.58 years (from 1 April 2003 to 31 October 2006) and then divided by the amount of money kept in the SAFEX account for margin.

<sup>98</sup> Please refer to the accompanying CD for a complete account of all the trades for each of the individual pairs.

TABLE 7.2 FINAL RESULTS: RSI

Pair	Total number of trades	Number of successful trades	Percentage of successful trades	Brokerage paid to retain positions	Profit or Loss in S.A. rand/ton
WMSPOT & WM 1 away	45	24	53.33%	-43.50	33.50
WMSPOT & YMSPOT	44	26	59.09%	-60.00	-31.00
WMSPOT & YM 1 away	35	23	65.71%	-43.50	128.50
WM 1 away & WM 2 away	29	13	44.83%	-27.00	-150.00
WM 1 away & YMSPOT	45	27	60.00%	-43.50	277.50
WM 1 away & YM 1 away	43	29	67.44%	-27.00	203.00
WM 1 away & YM 2 away	42	23	54.76%	-27.00	175.00
WM 2 away & YM 1 away	58	34	58.62%	-27.00	792.00
WM 2 away & YM 2 away	49	28	57.14%	-27.00	145.00
WM 3 away & YM 3 away	51	39	76.47%	-27.00	830.00
YMSPOT & YM 1 away	44	27	61.36%	-43.50	194.50
YM 1 away & YM 2 away	29	16	55.17%	-43.50	169.00
Totals: All Pairs	514	309	60.12%	-439.50	2767.00

When making use of the RSI to determine entry and exit points, more trades were performed over the test period. Of the 514 trades that were performed, 309 resulted in a profit (60.12 percent). Only the WMSPOT & YMSPOT and WM 1 away & WM 2 away pairs realised losses over the test period. The pair that earned the highest profit was WM 3 away & YM 3 away (R 830,00/tonne) followed by WM 2 away & YM 1 away (R 792,00/tonne).

A total profit of R 2 767,00/tonne was realised over the test period. This result is a considerable improvement on the loss realised when taking all the indicators into consideration. If the same example of ten contracts were traded for each pair by only taking cognisance of the signals given by the RSI, a total profit of R 2 767 000,00 would have been realised.

As before, to realise this profit R 1.76 million was permanently kept in a SAFEX margin account. In per annum terms, R 1.76 million had to be kept in the SAFEX margin account in order to realise a profit of R 772 905,03. Calculated as a rate of return on equity per annum, this strategy renders a return of 43.92 percent on the initial capital kept in the SAFEX margin account. This result is far better than the previous result, not only with regard to the amount of profit earned over the test period, but also in terms of the amount of successful trades performed over the same period.

Making use of only the RSI as a tool to identify entry and exit points is an improvement on the previous result. To further assess the profitability of pairs trading for the agricultural commodities trader, the Stochastic oscillator was tested next. These results are discussed in section 7.4.3.2 .

#### 7.4.3.2 Utilising the Stochastic Oscillator to Identify Entry and Exit Points

As before, the profit or loss made over the test period for the Stochastic oscillator tests are expressed as a rand per metric tonne value. The results reported in Table 7.3 were obtained by relying on the Stochastic oscillator alone to identify entry and exit points.<sup>99</sup>

TABLE 7.3 FINAL RESULTS: THE STOCHASTIC OSCILLATOR

Pair	Total number of trades	Number of successful trades	Percentage of successful trades	Brokerage paid to retain positions	Profit or Loss in S.A. rand/ton
WMSPOT & WM 1 away	64	34	53.13%	-43.50	42.50
WMSPOT & YMSPOT	70	38	54.29%	-60.00	104.00
WMSPOT & YM 1 away	72	39	54.17%	-43.50	329.50
WM 1 away & WM 2 away	77	12	15.58%	-27.00	-485.00
WM 1 away & YMSPOT	66	29	43.94%	-43.50	9.50
WM 1 away & YM 1 away	76	38	50.00%	-27.00	309.00
WM 1 away & YM 2 away	64	30	46.88%	-27.00	-2.00
WM 2 away & YM 1 away	77	37	48.05%	-27.00	264.00
WM 2 away & YM 2 away	68	32	47.06%	-27.00	-206.00
WM 3 away & YM 3 away	72	27	37.50%	-27.00	-526.00
YMSPOT & YM 1 away	78	36	46.15%	-43.50	7.50
YM 1 away & YM 2 away	86	22	25.58%	-43.50	-598.00
Totals: All Pairs	870	374	42.99%	-439.50	-751.00

Compared to the results of the tests performed by taking all the indicators into consideration, this strategy was more successful in monetary terms, only realising a loss of R 751,00/tonne. When comparing this strategy's success in terms of the amount of profitable trades, this strategy fared worse than the RSI results, only managing to realise a profit 42.99 percent of the time.

Even though only five of the twelve pairs realised a loss over the test period, only 374 of the 870 trades ended up in a profit. The pair realising the biggest profit for this period

<sup>99</sup> Please refer to the accompanying CD for the complete set of results regarding all the trades for each of the individual pairs performed when making use of only the Stochastic oscillator.

was WMSPOT & YM 1 away (R 329,50/tonne), followed by WM 1 away & YM 1 away (R 309,00/tonne). The pair that realised the biggest loss for the same period was YM 1 away & YM 2 away (–R 598,00/tonne) followed by WM 3 away & YM 3 away (–R 526,00/tonne).

Extending this loss for ten contracts traded for each pair by only taking cognisance of the signals given by the Stochastic oscillator, a total loss of R 751 000,00 would have been realised. As before, keeping R 1.76 million in a SAFEX margin account, a loss of –R 209 776,54 are realised per annum.

As a rate of return on equity per annum, making use of the Stochastic oscillator alone, provides a loss of 11.92 percent on the initial capital kept in the SAFEX margin account. The use of this technical analysis tool does not provide enough incentive to take such risks. To further assess the viability of a pairs trading strategy for the agricultural commodities trader, Bollinger bands were employed next. These results are discussed in section 7.4.3.3.

#### 7.4.3.3 Utilising Bollinger Bands to Identify Entry and Exit Points

Table 7.4 below reports on results of the tests performed by making use of Bollinger bands to identify entry and exit points. As was the case with the previous reports, the far right-hand side column reports the amount of money that was made or lost for each of the twelve pairs over the test period.<sup>100</sup>

---

<sup>100</sup> Please refer to the accompanying CD for the complete record of all the trades for each of the individual pairs in making use of Bollinger bands.

TABLE 7.4 FINAL RESULTS: BOLLINGER BANDS

Pair	Total number of trades	Number of successful trades	Percentage of successful trades	Brokerage paid to retain positions	Profit or Loss in S.A. rand/ton
WMSPOT & WM 1 away	30	23	76.67%	-43.50	237.50
WMSPOT & YMSPOT	20	14	70.00%	-60.00	278.00
WMSPOT & YM 1 away	29	21	72.41%	-43.50	377.50
WM 1 away & WM 2 away	26	15	57.69%	-27.00	222.00
WM 1 away & YMSPOT	27	19	70.37%	-43.50	111.50
WM 1 away & YM 1 away	30	18	60.00%	-27.00	465.00
WM 1 away & YM 2 away	28	18	64.29%	-27.00	193.00
WM 2 away & YM 1 away	27	22	81.48%	-27.00	618.00
WM 2 away & YM 2 away	29	22	75.86%	-27.00	686.00
WM 3 away & YM 3 away	25	15	60.00%	-27.00	572.00
YMSPOT & YM 1 away	27	22	81.48%	-43.50	533.50
YM 1 away & YM 2 away	16	8	50.00%	-43.50	120.00
Totals: All Pairs	314	217	69.11%	-439.50	4414.00

When comparing the Bollinger bands results to the previous results, this strategy was the most successful yet. As can be seen in Table 7.4, all of the pairs realised a profit over the test period. This strategy also required less trades than any of the other technical analysis tests. Note that, 217 of the 314 trades, or 69.11 percent of all trades, resulted in a profit. It is furthermore noteworthy that, with the exception of two pairs, all of the pairs recorded profitable trades more than 60 percent of the time. Even though some of the amounts are small, the agricultural commodities trader can be confident that he will earn a return on his investment.

In terms of the entire portfolio, a total profit of R 4 414,00/tonne was realised for the test period. The pair realising the biggest profit for this period was WM 2 away & YM 2 away (R 686,00/tonne), followed by WM 2 away & YM 1 away (R 618,00/tonne). The pair that realised the smallest profit for the same period was WM 1 away & YMSPOT (R 111,50/tonne) followed by YM 1 away & YM 2 away (R 120,00/tonne).

If ten contracts were traded for each pair, only making use of Bollinger bands to indicate entry and exit points for each trade, a total profit of R 4 414 000,00 would have been realised over the test period. In per annum terms, by keeping R 1.76 million in the SAFEX margin account, a profit of R 1 232 960,89 per year was generated.

This amount of profit translates to a rate of return on equity per annum of 70.05 percent. It is important to note that this return compares favourably with returns earned by other high risk investments. It is thus clear that a pairs trading strategy can be a viable option for the agricultural commodities trader when making use of Bollinger bands alone. The next assessment regarding the viability of a pairs trading strategy for the agricultural commodities trader follows in section 7.4.3.4. MACD is discussed here as a tool for identifying entry and exit points.

#### 7.4.3.4 Utilising MACD to Identify Entry and Exit Points

MACD was the second last technical analysis tool to be tested for its ability to identify entry and exit points. Table 7.5 below reports these results, with the far right-hand side column reporting the amount of money that was made or lost for each of the twelve pairs over the test period.<sup>101</sup>

TABLE 7.5 FINAL RESULTS: MACD

Pair	Total number of trades	Number of successful trades	Percentage of successful trades	Brokerage paid to retain positions	Profit or Loss in S.A. rand/ton
WMSPOT & WM 1 away	130	26	20.00%	-43.50	-915.50
WMSPOT & YMSPOT	118	34	28.81%	-60.00	-660.00
WMSPOT & YM 1 away	103	34	33.01%	-43.50	-128.50
WM 1 away & WM 2 away	154	15	9.74%	-27.00	-1261.00
WM 1 away & YMSPOT	98	36	36.73%	-43.50	-157.50
WM 1 away & YM 1 away	111	34	30.63%	-27.00	8.00
WM 1 away & YM 2 away	109	32	29.36%	-27.00	-243.00
WM 2 away & YM 1 away	98	33	33.67%	-27.00	-253.00
WM 2 away & YM 2 away	117	27	23.08%	-27.00	-528.00
WM 3 away & YM 3 away	106	38	35.85%	-27.00	-18.00
YMSPOT & YM 1 away	106	29	27.36%	-43.50	-681.50
YM 1 away & YM 2 away	161	17	10.56%	-43.50	-1837.00
Totals: All Pairs	1411	355	25.16%	-439.50	-6675.00

Working with MACD has two problems. The first is that a large number of entry and exit points are registered. This leads to an increase in the amount of money paid for brokerage. The second, and more critical problem, is that MACD does not register entry and exit points in time, resulting in the trades being entered at the wrong time.

<sup>101</sup> Please refer to the accompanying CD for the entire record of all the trades performed for each of the individual pairs in making use of MACD.

These problems are evident in the results. All but one of the pairs registered a loss, with eleven of the twelve pairs ending up with a loss. In terms of the entire portfolio, a total loss of R 6 675,00/tonne was achieved for the test period. The only pair realising a profit for this period was WM 1 away & YM 1 away (R 8,00/tonne). The pair realising the biggest loss was YM 1 away & YM 2 away (–R 1 837,00/tonne) followed by WM 1 away & WM 2 away (–R 1 261,00/tonne).

In extending this result to the example of ten contracts traded for each pair, a total loss of R 6 675 000,00 would have been realised over the test period, or R 1 864 525,14 per annum if MACD only was used to indicate entry and exit points for each trade. It is clear that MACD is not effective in identifying entry and exit points for this specific type of strategy, effectively realising a loss of 105.94 percent on equity.

#### 7.4.3.5 Utilising MA to Identify Entry and Exit Points

MA was the final technical analysis tool to be tested for its ability to identify entry and exit points, and the results are reported in Table 7.6 below. The far right-hand side column is reporting the amount of money that was made or lost for each of the twelve pairs over the test period.<sup>102</sup>

TABLE 7.6 FINAL RESULTS: MA

Pair	Total number of trades	Number of successful trades	Percentage of successful trades	Brokerage paid to retain positions	Profit or Loss in S.A. rand/ton
WMSPOT & WM 1 away	318	35	11.01%	-43.50	-2049.50
WMSPOT & YMSPOT	244	50	20.49%	-60.00	-953.00
WMSPOT & YM 1 away	272	56	20.59%	-43.50	-1047.50
WM 1 away & WM 2 away	339	19	5.60%	-27.00	-2233.00
WM 1 away & YMSPOT	249	60	24.10%	-43.50	-1311.00
WM 1 away & YM 1 away	266	58	21.80%	-27.00	-650.00
WM 1 away & YM 2 away	267	55	20.60%	-27.00	-2084.50
WM 2 away & YM 1 away	259	52	20.08%	-27.00	-779.00
WM 2 away & YM 2 away	265	50	18.87%	-27.00	-1647.00
WM 3 away & YM 3 away	246	58	23.58%	-27.00	-714.50
YMSPOT & YM 1 away	300	43	14.33%	-43.50	-2108.50
YM 1 away & YM 2 away	300	28	9.33%	-43.50	-2874.50
Totals: All Pairs	3325	564	16.96%	-439.50	-18452.00

<sup>102</sup> Please refer to the accompanying CD for the entire record of all the trades performed for each of the individual pairs in making use of MACD.

MA has the same two problems than MACD. A total of 3325 trades have been registered, making MA a very expensive tool regarding the amount of money paid for brokerage. Like MACD, MA also fails to register entry and exit points in time, resulting in the trades being entered into after highs and lows. The MA indicator realises a total loss of R 18 452,00/tonne, the worst of all the indicators used with all of the pairs registering a loss.

The biggest loss was realised by YM 1 away & YM 2 away (–R 2 874,50/tonne) followed by WM 1 away & WM 2 away (–R 2 233,00/tonne). The pair realising the smallest loss was WM 1 away & YM 1 away (–R 650,00/tonne), followed by WM 3 away & YM 3 away (–R 714,50/tonne). In the case where these results are extended to ten contracts traded for each pair, a total loss of R 18 452 000,00 would have been realised over the test period, if MACD only was used to indicate entry and exit points for each trade. MA should therefore not be used in identifying entry and exit points for this specific type of strategy.

#### 7.4.4 Conclusion

In testing the profitability of the final pairs portfolio selected in chapter 6, two approaches were followed. The first allowed for entry and exit points to be determined by making use of fundamental, technical, and seasonal analysis tools, while the second approach only allowed entry and exit points to be determined by each of the technical analysis tools individually.

The first approach lost money, and should not be considered as viable option when entering into this type of trading strategy. Mixed results were reported for the second approach. While the RSI and Bollinger bands fared well, the Stochastic oscillator, MACD and MA in particular performed poorly.

Accordingly, the Stochastic oscillator, MACD and MA should not be used as tools for identifying entry and exit points for this type of strategy. Conversely, the RSI and Bollinger bands do well enough for the agricultural commodities trader to consider a pairs trading strategy as a profitable trading option.

## 7.5 Summary

Before commencing with the test trading process, it was important to take note of several assumptions. These included assumptions regarding what entry and exit points are, and how these entry and exit points are identified by making use of fundamental analysis, the RSI, the Stochastic oscillator, Bollinger bands, MACD and MA's. Assumptions were also made about all costs, such as brokerage fees, margin payments, and opportunity costs, and the procedure followed in test trading the final pairs portfolio.

The test trading process was reported under two main approaches. The first involved test trading the final pairs portfolio by making use of all the indicators available to the trader. For this approach entry and exit points were registered once the fundamental and technical analysis tools were in accordance. This approach lost money, and only 28.91 percent of all the trades (for all the pairs over the test period) resulted in a profit. In terms of return on equity, the trader can expect to realise a loss of 127.28 percent on the initial capital in kept in his SAFEX margin account.

The second approach consisted of five tests, each performed by relying more heavily on one technical analysis tool's signals for entering and exiting the market at the time. The first test performed to this end, made use of the RSI as tool to identify entry and exit points. This indicator registered entry points when the 5-day RSI line crossed the 30 percent signal line from below. Similarly, exit points were registered as soon as the 5-day RSI line crossed the 70 percent signal line from above. By making use of the RSI in this manner, a total profit of R 2 767,00/tonne was realised over the test period. In terms of return per annum, a return of 43.92 percent was recorded on the initial equity amount.

The second technical analysis test employed the Stochastic oscillator as tool to identify entry and exit points. For this indicator to register an entry point, one of two scenarios must be true. In the first scenario, readings above 70 percent are considered overbought, while readings below 30 percent are considered oversold. In the second scenario, entry points were identified when the %D line crossed the %K line from above while exit points were registered when the %D line crossed the %K line from below.

The Stochastic oscillator does not render satisfactory results. Only 42.99 percent of all the trades resulted in a profit, rendering a loss on equity of 11.92 percent.

The third technical analysis test used to identify entry and exit points was Bollinger bands. Entry points were registered once the price crossed the bottom band line from below after breaking out of the two bands' range at the bottom. Equally, exit points were identified when prices crossed the top band line from above after breaking out of the band range at the top. By making use of this technical analysis tool, all of the pairs included in the final portfolio realised a profit for the test period. Besides its 100 percent record (all twelve pairs making a profit), 69.11 percent of all trades rendered a profit. The total amount of profit for the period of the study was R 4 414,00 resulting in a return on equity of 70.05 percent.

The fourth technical analysis test made use of MACD to identify entry and exit points for each trade. For MACD, entry points were registered once the MACD line rose above the signal line from below. In contrast, exit points were generated when the MACD line crossed the signal line from above. Using MACD to identify entry and exit points failed for two reasons: too many entry and exit points were identified, and signals to enter and exit the market were registered too late.

The fifth and final technical analysis test made use of MA to identify entry and exit points. For MA, entry points were registered once the actual price series crosses the MA line from below. In contrast, exit points were generated when the actual price series crossed the MA line from above. Like MACD, MA also failed to identify entry and exit points because too many entry and exit points were identified, and signals to enter and exit the market were registered after the high and low points.

A loss was recorded when making use of all the indicators used in combination, when making use of MACD, the Stochastic oscillator and MA. Only the RSI and Bollinger bands that rendered satisfactory results. When applying the RSI, a return on equity of 43.92 percent was realised over the test period, while a rate of return of 70.05 percent was realised when employing Bollinger bands to identify entry and exit points.

It is important to note that, even though these two technical analysis tools proved to be effective in the pairs trading process, there are a variety of different technical analysis and other tools that can be employed for identifying entry and exit points. Moreover, if more emphasis is placed on the fundamental factors during each trade, an even higher success rate might be achieved.

A profit can thus be realised even if the RSI and Bollinger bands are not used to indicate entry and exit points on condition that the pairs included in the agricultural commodities trader's portfolio retain the statistical relationship between them. However, if the relationship between pairs disintegrates, the trader will not be able to continue profiting from this strategy.

# CHAPTER 8

## Conclusion

### 8.1 Introduction

Chapter 8 commences with a brief review of the aim of this research project. Thereafter, the relevant information discussed in the previous chapters are outlined. Firstly, a brief overview is provided of fundamental analysis and the factors that influence agricultural commodity prices in section 8.3. Next, an overview on technical analysis is given in section 8.4. This is followed by a short review of ARMA and ARIMA, GARCH and UCM in section 8.5. Thereafter, overviews of testing for seasonality, pairs selection and trading agricultural commodity pairs are presented in section 8.6, section 8.7 and section 8.8, respectively. After this, conclusions of this thesis are drawn in section 8.9, and finally suggestions for further study are made in section 8.10.

### 8.2 Aim of the Research Project

The aim of the research project on which this thesis reports, was to determine whether pairs trading can be used as a profitable option for the agricultural commodities trader in South Africa. Although pairs trading has been used as a trading strategy by some agricultural commodities traders in South Africa, no formal process for implementing this strategy in the South African agricultural commodities market has been documented to date. This research project thus aimed to fill this gap by endeavouring to formulate and report on a systematic process for selecting an agricultural commodity pairs portfolio and for implementing this trading strategy successfully in the South African agricultural commodities market.

### 8.3 Study Review: Fundamental Analysis

Chapter 2 explored the fundamental factors influencing the prices of those agricultural commodities traded on SAFEX. Agricultural commodity prices are influenced by both the supply and demand of the commodity itself, as well as its complimentary and substitute products. The chapter summarised the history of these agricultural

commodities, and explained the supply and demand factors that influence the prices of these commodities. These factors must be taken into consideration by the pairs trader in making trading decisions.

## 8.4 Study Review: Technical Analysis

Chapter 3 discussed the technical analysis tools that were used in chapter 7 during the trading process. This chapter looked at several market strength and moving average indicators most frequently used by agricultural commodity traders. Five specific indicators were highlighted: the RSI, the Stochastic oscillator, Bollinger bands, MACD and MA. Together with the fundamental analysis covered in chapter 2, and seasonal measures covered in chapter 5, technical analysis tools are used by the trader to pin point entry and exit points for every trade.

## 8.5 Study Review: Measures used to test for Seasonality

Chapter 4 assessed the different models employed in testing for seasonal patterns in the data. Both deterministic and stochastic models were utilised to this end. The two deterministic models introduced in chapter 4 were the ARMA- and ARIMA-type models, and the general GARCH model. The stochastic UCM model was also introduced. All of these models were used in chapter 5 in testing for seasonal patterns in the data.

## 8.6 Study Review: Testing for Seasonality

In chapter 5, the maize series were tested for seasonal patterns by means of the tests introduced in chapter 4. The first analysis tested for seasonality by making use of several GARCH models. This model failed to identify any realistic seasonal pattern in the data. The second model tested for the presence of seasonal patterns in the data by employing the X12 ARIMA-type test. Here, identifiable seasonality was reported for the YMSPOT and YM 1 away series.

Even though identifiable seasonality was not reported for any of the other series, both the Henderson-curve and the seasonal factors series were successful in identifying the

seasonal patterns in the data. The tests for hidden seasonal patterns in the data by means of UCM were also successful in identifying the seasonal patterns in the data.

The results of the seasonality tests were utilised in two ways. In the first way, these results were used as guides to the seasonal patterns in the data, and should be used in conjunction with fundamental analysis by the pairs trader. In the second way, the X12 and UCM results were also used as additional pairs selection tools.

## 8.7 Study Review: Pairs Selection

Chapter 6 reported on the pairs selection process. This was performed under two main sections. The first section discussed several pairs selection tools found in the literature, including statistical measures such as co-integration, Pearson's correlation, Spearman's rank correlation and Beta. The second section reported on the results of the X12 and UCM tests performed in chapter 5.

The pairs selection process was executed in two phases: testing for the strength of the statistical relationship between the different maize series by means of co-integration, Pearson's correlation and Spearman's rank correlation; and testing for co-movement between the maize series by means of the correlation between their deterministic and stochastic seasonal patterns respectively.

During the first step of phase one in the pairs selection process, tests for co-integration were performed in order to establish whether the different maize series had any long-run statistical relationships between them. According to the co-integration results, twenty-eight preliminary pairs were selected. The second step of phase one served as cross-check to validate the preliminary pairs. To this end, tests were done to determine the Pearson correlation coefficients of the preliminary pairs. The final results indicated that all twenty-eight of the preliminary pairs, selected during the first step, had correlation coefficients of 0.9 and higher. The third step of phase one involved testing for the Spearman rank correlation coefficients between the twenty-eight preliminary pairs. Once more, the results of the Spearman rank test indicated that all of the preliminary pairs were to remain in the preliminary pairs portfolio, reporting correlation coefficients of 0.9 and higher.

During step one of the second phase, in the pairs selection process, tests were performed to find the correlation coefficients between the Henderson-curves of the preliminary pairs. The results indicated that all of the preliminary pairs were to be included in the final pairs portfolio. During step two of the second phase, the correlation coefficients between the various maize series' seasonal factors series were calculated, after which some of the original preliminary pairs were discarded from the portfolio. According to the results, only seventeen of the original twenty-eight pairs were selected to remain in the final pairs portfolio. During step three of the second phase, the correlation coefficients between the various maize series' hidden seasonal series were calculated.

Based on these results, the final pairs portfolio was reduced to include only twelve of the original twenty-eight pairs initially selected. The final pairs portfolio was comprised of the following pairs: WMSPOT & WM 1 away, WMSPOT & YMSPOT, WMSPOT & YM 1 away, WM 1 away & WM 2 away, WM 1 away & YMSPOT, WM 1 away & YM 1 away, WM 1 away & YM 2 away, WM 2 away & YM 1 away, WM 2 away & YM 2 away, WM 3 away & YM 3 away, YMSPOT & YM 1 away, and YM 1 away & YM 2 away.

## 8.8 Study Review: Trading Agricultural Commodity Pairs

Once all the pairs have been validated, the trader will be able to trade this final pairs portfolio with success. For the purposes of this project, two approaches were followed. The first approach used fundamental analysis and technical analysis tools in identifying entry and exit points for all the trades. This realised, a total loss of R 8019,50/tonne for the period 1 April 2003 to 31 October 2006.

In the second approach, each of the technical analysis tools was tested in isolation from each other. For this approach hidden seasonal patterns were ignored, and the trading signals of the individual indicators carried more weight than the fundamental analysis signals. Results were reported for four technical analysis tools: the RSI, the Stochastic oscillator, Bollinger bands, MACD and MA.

In using the RSI as tool to identify entry and exit points a total profit of R 2 767,00/tonne was realised for the period 1 April 2003 to 31 October 2006. In terms

of return per annum, a return on equity of 43.92 percent was recorded on the initial equity amount. This result compares favourably with the rates of return for investment strategies that carry the same amount of risk.

The results of the tests performed in making use of the Stochastic oscillator resulted in a loss. The pairs portfolio makes a loss of R 751,00 for period 1 April 2003 to 31 October 2006, with only 42.99 percent of all the trades resulted in a profit, rendering a loss on equity of 11.92 percent.

The best results were reported for the tests performed in making use of Bollinger bands to identify entry and exit points. Only 30.89 percent of all trades resulted in a loss, and a total profit of R 4 414,00 was reported for the period 1 April 2003 to 31 October 2006. This means that a rate of 70.05 percent was realised on the initial equity in the margin account.

Making use of MACD did not render satisfactory results. In using this technical analysis tool, a total loss of R 6 675,00 was reported for the period 1 April 2003 to 31 October 2006.

The worst results were reported for the tests performed with MA to identify entry and exit points. All of the trades resulted in a loss, and a total loss of R 18 452,00 was reported for the period 1 April 2003 to 31 October 2006.

Two important findings were made during the test trading process. The first is that UCM proved useful in identifying the underlying seasonal patterns in the data, and the second is that UCM works well as pairs selection tool. The second finding is that it is more profitable to isolate the use of the technical analysis tools for the purpose of identifying entry and exit points during the agricultural commodity pairs trading process in South Africa.

## 8.9 Conclusion

In considering the final results it is clear that it is indeed possible for the agricultural commodities trader in South Africa to profit from the implementation of a pairs trading

strategy. These results are, however, reliant on the condition that all the pairs selected during the pairs selection process retain the statistical relationship between them. This thesis has reported on a systematic process that can be followed to successfully identify different agricultural commodities with similar statistical relationships as pairs.

In order to guarantee that the initial pairs selection remains robust, it is imperative that the agricultural commodities trader tests for pairs on a regular basis. In doing so, he will be assured that the commodity series comprising each pair always revert back to a common mean. It is this mean-reverting property that makes pairs trading successful.

When the agricultural commodities trader is confident that all the pairs in his portfolio will continue to remain pairs in the future, he will be able to trade these pairs successfully by relying on the signals provided by both fundamental and technical analysis tools. It is important to note that, even though using the RSI and Bollinger bands was successful in identifying entry and exit points, the use of any other analysis tools may be successful as long as the statistical relationship between the agricultural commodities comprising the pairs holds.

## 8.10 Recommendations for Further Study

The knowledge gained in this research project can be used in the expansion of two areas. The first pertains to the pairs selection process itself, and the second to the extension of the application of both the pairs selection and pairs trading processes to a broader market.

The literature shows that the differencing process in achieving stationarity in economical time series data is sometimes a fraction. If it is found that the differencing process for the South African agricultural commodities data is a fraction between 0.5 and 1, it will mean that the data series is non-stationary, which implies that the process is mean reverting (Maddala and Kim, 2000:298). Since mean reversion is one of the key assumptions necessary for the success of a pairs trading strategy, Johansen's co-integration tests might be replaced with tests for fractional co-integration during the pairs selection process in refining the pairs selection process even further.

---

As this research project aimed to test for the viability of implementing a pairs trading strategy in the South African agricultural commodities market, a natural extension of this study would be to replicate the pairs selection process for finding agricultural commodity pairs between SAFEX and other markets, such as CBOT. By replicating the processes documented in this thesis it should be possible to search for, and successfully trade 'cross-market' pairs.

## REFERENCES

- ACHELIS, S.B. 2001. *Technical analysis from A to Z*. 2nd ed. New York, NY: McGraw-Hill. 380 p.
- AGGARWAL, R. & RIVOLI, P. 1989. Seasonal and day of the week effect in four emerging stock markets. *Financial review*, 24(4):541-550.
- AKGIRAY, V. 1989. Conditional heteroskedasticity in time series of stock returns: evidence and forecasts. *Journal of business*, 62(1):55-80, Jan.
- AL-DEEHANI, T.M. 2006. Seasonality as an unobservable component: the case of Kuwait Stock Market. *Applied financial economics*, 16(6):471-478, Mar.
- ALEXANDER, C.O. & DIMITRIU, A. 2002. The co-integration alpha: enhanced index tracking and long-short equity market neutral strategies. *ISMA Discussion Papers in Finance*, paper 8, April. <http://www.ismacentre.rdg.ac.uk/pdf/discussion/DP2002-08.pdf> Date of access: 11 Aug. 2006.
- AL-SAAD, K., & MOOSA, I. 2005. Seasonality in Stock Returns: evidence from an emerging market. *Applied financial economics*, 15(1):63-71, Jan.
- ATHANASSAKOS, G. & ROBINSON, M.J. 1994. The day-of-the-week anomaly: the Toronto Stock Exchange experience. *Journal of business finance and accounting*, 21(6):833-856.
- AUSTRALIAN BUREAU METEOROLOGY (ABM). 2006. Southern Oscillation Index. <http://www.bom.gov.au/lam/glossary/soid.htm> Date of access: 7 Aug. 2006.
- BAILLIE, R.T. & DEGENNARO, R.P. 1990. Stock returns and volatility. *Journal of financial and quantitative analysis*, 25(2):203-214, Jun.
- BERGES, A., McCONNELL, J. & SCHLARBAUM, G. 1984. An investigation of the turn-of-the-year effect, the small firm effect and the tax-loss selling pressure hypothesis in Canadian stock returns. *Journal of finance*. 39:185-192, Mar.
- BERNSTEIN, J. 2000. *How the futures market work*. 2nd ed. New York, NY: Prentice-Hall. 253 p.
- BERUMENT, H. & KIYMAZ, H. 2001. The day of the week effect on stock market volatility. *Journal of economics and finance*, 25(2):181-193, Summer.
- BERUMENT, H. & KIYMAZ, H. 2003. The day of the week effect on stock market volatility and volume: international evidence. *Review of financial economics*, 12:363-380, Jun.
- BLACK, F. 1976. Studies of stock price volatility changes. (*In the Proceedings of 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*. p. 177-181.)

- BOLLERSLEV, T., ENGLE, R. & WOOLDRIDGE, J.M. 1988. A capital asset pricing model with time varying covariances. *Journal of political economy*, 96(1):116-131, Feb.
- BROWN, P., KEIM, D.B., KELEIDON, A.W. & MARSH, T.A. 1983. Stock return seasonalities and the tax-loss-selling-hypothesis: analysis of the arguments and Australian evidence. *Journal of financial economics*, 12:105-127, Jun.
- BURKE, S.P. & HUNTER, J. 2005. Modelling non-stationary economic time series: a multivariate approach. New York, NY: Palgrave MacMillan. 253 p.
- CAMPBELL, J.Y. & HENTSCHEL, L. 1992. No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of financial economics*, 31:281-318.
- CASTLEMAN, S.D. 2003. Pair trading with implied volatility. *Futures: news, analysis & strategies for futures, options & derivatives traders*, 32(14):38-41, Nov.
- CHANDE, T.S. 2001. Beyond technical analysis: How to develop and implement a winning trading system. 2nd ed. Hoboken, NJ: Wiley. 400 p.
- CHANG, T., NIEH, M. & WEI, C. 2006. Analysis of long-run benefits from international equity diversification between Taiwan and its major trading partners: an empirical note. *Applied economics*, 38:2277-2283.
- CHANG, E., PINEGAR, M. & RAVICHANDRAN, R. 1993. International evidence on the robustness of the day of the week effect. *Journal of financial and quantitative analysis*, 28(4):497-513, Dec.
- CHATFIELD, C. 2004. The analysis of time series: an introduction. 6th ed. London: Chapman & Hall. 333 p.
- CHICAGO BOARD OF TRADE (CBOT). 2006a. Our history. <http://www.cbot.com/cbot/pub/page/0,3181,942,00.html> Date of access: 8 Aug. 2006.
- CHICAGO BOARD OF TRADE (CBOT). 2006b. Side-by-side agricultural trading. <http://www.cbot.com/cbot/pub/page/0,3181,963,00.html> Date of access: 9 Aug. 2006.
- CHICAGO MERCANTILE EXCHANGE (CME). 2006. History of innovation. <http://www.cme.com/about/ins/caag/history2801.html> Date of access: 8 Aug. 2006.
- COLBY, R.W. 2002. The encyclopaedia of technical market indicators. 2nd ed. New York, NY: McGraw-Hill. 832 p.
- CORHAY, A., FATEMI, A. & RAD, A.T. 1995. On the presence of a day-of-the-week effect in the foreign exchange market. *Managerial finance*, 21:32-43.
- CORNELL, B. 1985. The weekly patterns in stock returns cash versus futures: a note. *The journal of finance*, 40(2):583-588, Jun.

- CROOME, S. 2003. Understanding volatility measurements. [http://www.tradetrek.com/education/risk\\_management.asp](http://www.tradetrek.com/education/risk_management.asp) Date of access: 11 April 2006.
- CROSS, F. 1973. The behaviour of stock prices on Friday and Monday. *Financial analysts journal*, 31(6):67-69.
- DAGUM, E.B. 1980. The X-11-ARIMA Seasonal Adjustment Method. Catalogue 12(564E), Ottawa: Statistics Canada.
- DEFUSCO, R.A., MCLEAVEY, D.W., PINTO, J.E. & RUNKLE, D.E. 2007a. Correlation and regression. (In Chartered Financial Analyst (CFA) Institute. Ethical and professional standards and quantitative methods: CFA programme curriculum, level 1, vol. 1. Boston, MA: Pearson. p. 497-557).
- DEFUSCO, R.A., MCLEAVEY, D.W., PINTO, J.E. & RUNKLE, D.E. 2007b. Hypothesis testing. (In Chartered Financial Analyst (CFA) Institute. Ethical and professional standards and quantitative methods: CFA programme curriculum, level 1, vol. 1. Boston, MA: Pearson. p. 451-496).
- DE SANTIS, G. & IMROHOROGLU, S. 1997. Stock returns and volatility in emerging financial markets. *Journal of international money and finance*, 16:561-579.
- DUBOIS, M.L. 1986. The day-of-the-week effect: the international evidence. *Journal of banking and finance*, 20:1463-1484.
- DYL, E. & MABERLY, E.D. 1986. The anomaly that isn't there: a comment on Friday the thirteenth. *Journal of finance*, 43(5):1285-1286, Dec.
- EHRMAN, D.S. 2004. Pairs Trading: new look at an old strategy. *Futures: news, analysis & strategies for futures, options & derivatives traders*, 33(6):32-34, May.
- EHRMAN, D.S. 2006. The handbook of pairs trading: strategies using equities, options, & futures. Hoboken, NJ: Wiley. 258 p.
- EURONEXT. 2006. History: Euronext.liffe. [http://www.euronext.com/editorial/wide/0,5371,1732\\_2087568,00.html](http://www.euronext.com/editorial/wide/0,5371,1732_2087568,00.html) Date of access: 9 Aug. 2006.
- ENGLE, R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987-1007, Jul.
- ENGLE, R.F. 2001. GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of economic perspectives*, 15(4):157-168, Autumn.
- ENGLE, R.F. & BOLLERSLEV, T. 1986a. Modelling the persistence of conditional variances. *Economic reviews*, 5:1-50.
- ENGLE, R.F. & GRANGER, C.W.D. 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251-276.

- ENGLE, R.F., LILIEN, D.M. & ROBINS, R.P. 1987. Estimating time risk premia in the term structure: the ARCH-M model. *Econometrica*, 55(2):391-407, Mar.
- ETZKORN, M. 1997. Trading with oscillators: pinpointing market extremes – theory and practice. New York, NJ: Wiley. 176 p.
- FLANNERY, M.J. & PROTOPAPADAKIS, A.A. 1988. From T-bills to common stocks: investigating the generality of intra-week return seasonality. *The journal of finance*, 43(2):431-450, Jun.
- FRENCH, K. 1980. Stock returns and the weekend effect. *Journal of financial economics*, 8(1):55-69.
- FRENCH, K., SCHWERT, G. & STAMBAUGH, R.F. 1987. Expected stock returns and volatility. *Journal of financial economics*, 19:3-29.
- FTSE INDEX COMPANY (FTSE). 2005. FTSE guide to hedge funds. [http://www.ftse4good.com/indices\\_marketdata/ftsehedge/hedgeguide.pdf](http://www.ftse4good.com/indices_marketdata/ftsehedge/hedgeguide.pdf) Date of access: 24 April 2006.
- GARDNER, M. J. & MILLS, D. L. 1994. Managing financial institutions: an asset liability approach 3rd.ed. Orlando: Dryden Press. 980 p.
- GATEV, E.G., GOETZMANN, W.N. & ROUWENHORST, K.G. 1999. Pairs trading: performance of a relative value arbitrage rule. Working Paper 7032, Mar. <http://www.nber.org/papers/w7032> Date of access: 18 Mar. 2006.
- GAY, G. & KIM, T. 1987. An investigation into seasonality in the futures market. *Journal of futures market*, 7:169-181.
- GELDENHUYS, T. 2006. Verbal communication with the author. Lichtenburg. (Recording in the possession of the Noordwes Koöperasie.)
- GEMAN, H. 2005. Commodities and Commodity Derivatives. Hoboken, NJ: Wiley. 396 p.
- GEYSER, M. & CUTTS, M. 2007. SAFEX maize price volatility scrutinised. *Agrekon*, 46(3):291-305 Sept.
- GIBBONS, M. & HESS, P. 1981. Day of the week effects and asset returns. *Journal of business*, 54(4):579-596, Oct.
- GITAM INSTITUTE OF FOREIGN TRADE (GIFT). 2005. Indicator analysis for commodity derivatives. <http://www.indiainfoline.com/bisc/ari/comm.pdf> Date of access: 20 Apr. 2006.
- GOMEZ, V. & MARAVALL, A. 1996. Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User. Working Paper 9628.

- GOODWORTH, T.R.J. & JONES, C.M. 2004. Building a risk measurement framework for hedge funds and funds of funds. [http://www.jbs.cam.ac.uk/research/working\\_papers/2004/wp0408.pdf](http://www.jbs.cam.ac.uk/research/working_papers/2004/wp0408.pdf) Date of access: 17 Apr. 2006.
- GRANGER, C.W.J. 1981. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121-130.
- GRANGER, C.W.J. 1986. Developments in the study of co-integrated economic variables. *Oxford bulletin of economics and statistics*, 48:213-228.
- GUJARATI, D.N. 2003. Basic econometrics. 4th ed. New York, NY: McGraw-Hill. 982 p.
- HAGSTORM, R.G. 2005. The Warren Buffet way. 2nd ed. Hoboken, NJ: Wiley & Sons. 245 p.
- HAMAOKA, Y., MASULIS, R.W. & NG, V. 1990. Correlations in price changes and volatility across international stock markets. *The review of financial studies*, 3(2):281-307.
- HANSEN, J.W., JONES, J.W., KIKER, C.F. & HODGES, A.W. 1999. El Niño — Southern Oscillation impacts on winter vegetable production in Florida. *Journal of climate*, 12:92-102, Jan.
- HARVEY, A.C. 1990. Forecasting, structural time series models and the Kalman Filter. Cambridge: Cambridge University Press. 554 p.
- HARVEY, A.C. 1997. Trends, cycles and autoregressions. *The economic journal*, 107(440):192-201, Jan.
- HARVEY, A.C. & SCOTT, A. 1994. Seasonality in dynamic regression models. *The economic journal*, 104(427):1324-1345, Nov.
- HARVEY, A.C. & TRIMBUR, T.M. 2003. Generalised model-based filters for extracting cycles and trends in economic time series. *The Review of Economics and Statistics*, 85(2):244-255, Mar.
- HARVEY, A.C., TRIMBUR, T.M. & VAN DIJK, H.K. 2006. Trends and cycles in economic time series: a Bayesian approach. *Journal of econometrics*, 140:618-649, Aug. Available: ScienceDirect. Date of access: 5 Dec. 2007.
- HATANAKA, M. 1996. Time series-based econometrics: unit roots and co-integration. New York, NY: Oxford University Press. 294 p.
- HEIM, R.R., LAWRIMORE, J.H., WUERTZ, D.B., WAPLE, A.M. & WALLIS, T.W.R. 2003. The REDTI and MSI: two new national climate impact indices?. *Journal of applied meteorology*, 42:1435-1442.

- HEISER, C. B. 1981. *Seed to civilisation*. 2nd ed. San Francisco, CA: Freeman. 254 p.
- HEYMANS, A. 2005. The day-of-the-week effect as a risk for hedge fund managers. Potchefstroom: NWU. (Dissertation – M.Com) 132 p.
- HINEBAUGH, M. 1985. Wheat. (In Pring, M.J., ed. *The McGraw-Hill handbook of commodities and futures*. New York, NY: McGraw-Hill. § 6-3–6-34).
- HINEBAUGH, M. 1985. Corn. (In Pring, M.J., ed. *The McGraw-Hill handbook of commodities and futures*. New York, NY: McGraw-Hill. § 7-1–7-41).
- HINEBAUGH, M. 1985. Soybeans. (In Pring, M.J., ed. *The McGraw-Hill handbook of commodities and futures*. New York, NY: McGraw-Hill. § 8-1–8-32).
- HIRT, G.A. & BLOCK, S.B. 1994. *Fundamentals of investment management*. 4th ed. Boston, MA: Irwin. 787 p.
- HOERLING, M.P., KUMAR, A., & ZHONG, M. 1997. El Nino, La Nina, and the nonlinearity of their teleconnections. *J. Clim*, 10:1769–1786.
- HULL, J.C. 2000. *Options futures and other derivatives*. 4th ed. London: Prentice-Hall. 698 p.
- HUNTINGTON, E.V. 1919. Mathematics and statistics, with an elementary account of the correlation coefficient and the correlation ratio. *The American mathematical monthly*, 26(10):421-435, Dec.
- ISLAM, M.F. & HASAN, M.S. 2006. The Monetary model of the Dollar-Yen exchange rate determination: a co-integration approach. *International journal of business and economics*, 5(2):129-145.
- JAFFE, J. & WESTERFIELD, R. 1985a. Patterns in Japanese common stock returns: day of the week and turn of the year effects. *Journal of financial and quantitative analysis*, 20:261-272.
- JAFFE, J. & WESTERFIELD, R. 1985b. The week-end effect in common stock returns: the international evidence. *The journal of finance*, 40(2):433-454, Jun.
- JOHANSEN, S. 1988. Statistical analysis of co-integration vectors. *Journal of economic dynamics and control*, 12:231-254.
- JOHANSEN, S. 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6):1551-1580, Nov.
- JOHANSEN, S. 1995a. A statistical analysis of co-integration for I(2) variables. *Econometric theory*, 11(1):25–59.

- JOHANSEN, S. & JUSELIUS, K. 1990. Maximum likelihood estimation and inference on co-integration — with applications to the demand for money. *Oxford bulletin of economics and statistics*, 52(2):169-210.
- JSE SECURITIES EXCHANGE (SAFEX AGRICULTURAL PRODUCTS DIVISION). 2007. <http://www.safex.co.za> Date of access: 15 Aug. 2007.
- KATO, K. & SCHALLHEIM, J.S. 1985. Seasonal and size anomalies in the Japanese Stock Market. *Journal of financial and quantitative analysis*, 20(2):243-60, Jun.
- KEIM, D.B. 1983. Size related anomalies and stock return seasonality. *Journal of financial economics*, 12:13-32, Jun.
- KEIM, D.B. & STAMBAUGH, R.F. 1984. A further investigation of weekend effects in stock returns. *The journal of finance*, 39(3):819-840, Jul.
- KEYNES, J.M. 1936. The general theory of employment, interest and money. London: MacMillan. 403 p.
- KLEINMAN, G. 2001. Commodity futures and options: a step-by-step guide to successful trading. London: Prentice-Hall. 423 p.
- KOLB, R.W. 1997. Understanding futures markets. 5th ed. Malden, MA: Blackwell. 530 p.
- KOOP, G. 2005. Analysis of economic data. 2nd ed. Chichester: Wiley. 241 p.
- KOOPMAN, S.J., HARVEY, A.C., DOORNIK, J.A. & SHEPHARD, N. 2006. Structural time series analyser, modeller and predictor. London: Timberlake Consultants Ltd. 180 p.
- KOUTSOYIANNIS, A. 1977. Theory of Econometrics. 2nd ed. London: MacMillan. 681 p.
- KPMG. 1997. VAR: understanding and applying Value at Risk. London. Risk Publications. 393 p.
- KRUGEL, L.J. 2003. White maize futures contracts in South Africa. Potchefstroom: Potchefstroom University for Christian Higher Education. (Thesis – M.Com.) 93 p.
- KUNGLIGA VETENSKAPSAKADEMIEN (KVA). 2003. Time-series econometrics: co-integration and autoregressive conditional heteroskedasticity. Stockholm: KVA. 31 p. <http://nobelprize.org/economics/laureates/2003/ecoadv.pdf> Date of access: 20 Jul. 2005.
- LAKONISHOK, J. & LEVI, M. 1982. Weekend effect in stock return: a note. *Journal of finance*, 37(3):883-889, Jun.
- LEE, C.F., LEE, J.C., & LEE, A.C. 2000. Statistics for business and financial economics. 2nd ed. London: World Scientific. 976 p.

- LEIBTAG, E. 2008. Corn prices near record high, but what about food costs? *Amber waves*, 6(1):10-15, Feb.  
<http://www.ers.usda.gov/AmberWaves/February08/Features/CornPrices.htm> Date of access: 08 Mar. 2008.
- LEWIS, M. 1989. Stock market anomalies: a re-assessment based on the U.K. evidence. *Journal of banking and finance*, 13:675-696.
- LEVY, H. 2002. Fundamentals of investments. London: Pearson Education. 541 p.
- LITTERMAN, B. 2003. Modern investment management. Hoboken, NJ: Wiley. 593 p.
- LJUNG, G.M. & BOX, G.E.P. 1978. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297-303, Jan.
- MADDALA, G.S. 2002. Introduction to econometrics. 3rd ed. West Sussex: Wiley. 603 p.
- MADDALA, G.S. & KIM, I.M. 2000. Unit roots, co-integration, and structural change. Cambridge: Cambridge University Press. 505 p.
- MALAN, D. 2007. Verbal communication with the author. Riebeeckstad. (Recording in possession of the author.)
- MARTIN, R.V., WASHINGTON, R. & DOWNING, T.E. 2000. Seasonal maize forecasting for South Africa and Zimbabwe derived from an agroclimatological model. *Journal of applied meteorology*, 39:1473-1479, Sept.
- MASONSON, L.N. 2004. All about market timing. New York, NY: McGraw-Hill Professional. 288 p.
- MCEWAN, R. 2003. A simplified approach to pairs trading. *Futures: news, analysis & strategies for futures, options & derivatives traders*, 32(15):34-37, Dec.
- MCHATTIE, A. 1996. The investor's guide to warrants: capitalise on the fastest growing sector on the stock market. 2nd ed. London: Pitman Publishing. 334 p.
- MCMILLAN, D.G. 2005. Co-integrating behaviour between spot and forward exchange rates. *Applied financial economics*, 15:1135-1144.
- MEYERS, T.A. 1994. The technical analysis course: a winning program for investors and traders. New York, NY: McGraw-Hill. 311 p.
- MILLS, T.C., SIRIOPOULOS, C., MARKELLOS, R.N. & HARIZANIS, D. 2000. Seasonality in the Athens Stock Exchange. *Applied financial economics*, 10:137-142.
- MICROSOFT CORPORATION (MS). 2007. Microsoft® Office Excel® part of Microsoft® Office Enterprise 2007.

- MUKHERJEE, C., WHITE, H. & WUYTS, M. 1998. *Econometric and data analysis for developing countries*. New York, NY: Routledge. 496 p.
- MURPHY, J.J. 1986. *Technical analysis of the futures markets: a comprehensive guide to trading methods and applications*. New York, NY: New York Institute of Finance. 556 p.
- MURPHY, J.J. 1996. *The visual investor: how to spot market trends*. New York, NY: Wiley. 306 p.
- MURPHY, J.J. 1999. *Technical analysis of the financial markets: a comprehensive guide to trading methods and applications*. New York, NY: New York Institute of Finance. 542 p.
- NATIONAL CROP ESTIMATE COMMITTEE (NCEC). 2007. Area planted and sixth production forecast of summer crops: 2006/07 season. <http://www.nda.agric.za/> Date of access: 28 Aug. 2007.
- NELSON, D.B. 1990a. Stationarity and persistence in the GARCH(1,1) model. *Economic theory*, 6:318-334.
- NELSON, D.B. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2):347-370, Mar.
- NEW YORK BOARD OF TRADE (NYBOT). 2006. About NYBOT. <http://www.nybot.com/> Date of access: 9 Aug. 2006.
- OSBORNE, M.F.M., 1959. Brownian motion in the stock market. *Operations research*, 7:145-173, Mar.
- OGUM G., NOUYRIGAT G. & BEER F. 2002. An empirical analysis of Kenyan daily returns using EGARCH models. *Journal of research by CERAG*, 21:20 p. [http://www.cerag.org/axe\\_publicis.php?Caxe=1](http://www.cerag.org/axe_publicis.php?Caxe=1) Date of access: 14 Sep. 2005.
- OXMETRICS, 2006. STAMP™ Package. Version 4.04.
- PAFKA, S. & KONDOR, I. 2001. Evaluating the riskmetrics methodology in measuring volatility and Value-at-Risk in financial markets. <http://colbud.hu/pdf/Kondor/riskm.pdf> Date of access: 26 Mar. 2006.
- PANDEY, I.M. & CHOTIGEAT, T. 2005. Seasonality in Asia's emerging markets: India and Malaysia. (*In International Trade and Finance Association Conference Papers (Working Paper 53, Fall): the Proceedings of the 15th International Conference of the International Trade and Finance Association, Istanbul, Turkey, May 18-21. Berkeley, CA: Berkeley Electronic Press.*) <http://services.bepress.com/itfa/15th/art53> Date of access: 6 Dec. 2007.
- PHILLIPS, P.C.B. & PERRON, P. 1988. Testing for a unit root in time series regression. *Biometrika*, 75:335-346.

- PRESTON, T. 2005. Pairs trading. *Traders magazine: strategies*:40-44, Nov.  
<http://www.mediaserver.thinkorswim.com/%20articles/TPPairsTradingArticle.pdf> Date of access: 20 Jul. 2006.
- PRING, M.J. 2002. *Technical analysis explained: The successful investor's guide to spotting investment trends and turning points*. 4th ed. New York, NY: McGraw-Hill Professional. 560 p.
- PUTNAM, D.H., OPLINGER, E.S., HICKS, D.R., DURGAN, B.R. NOETZEL, D.M., MERONUCK, R.A., DOLL, J.D. & SCHULTE, E.E. 1990. Sunflower.  
<http://www.hort.purdue.edu/newcrop/afcm/sunflower.html> Date of access: 5 Sept. 2006.
- QUANTITATIVE MICRO SOFTWARE (QMS). 2004a. *EViews 5 Users Guide*. 978 p.
- QUANTITATIVE MICRO SOFTWARE (QMS). 2004b. *EViews. Version 5.1*.
- RAMANATHAN, R. 2002. *Introductory econometrics with applications*. 5th ed. New York, NY: Harcourt. 688 p.
- REINGANUM, M.R. 1983. The anomalous stock market behaviour of small firms in January: empirical tests for year-end tax effect. *Journal of financial economics*, 12:89-104, Jun.
- ROGALSKI, R.J. 1984. New findings regarding day of the week returns over trading and non-trading periods: a note. *Journal of finance*, 39(5):1603-1614, Dec.
- ROGERS, K. 1999. *An introduction to technical analysis*. New York, NY: Wiley. 182 p.
- ROSSOUW, W. 2006. Verbal communication with the author. Riebeeckstad. (Recording in possession of the author.)
- SOUTH AFRICAN GRAIN INFORMATION SERVICE (SAGIS). 2007. World maize production. <http://www.sagis.org.za/> Date of access: 29 Aug. 2007.
- SALVADOR, R.J. 1997. Maize. (In Werner, M.S., ed. *The encyclopaedia of Mexico: history, society and culture*, vol. 2. Chicago, ILL: Fitzroy Dearborn.)  
<http://www.questia.com/PM.qst?a=o&d=98882479> Date of access: 4 Jul. 2006.
- SAS INSTITUTE INC (SAS). 2006a. SAS/ETS<sup>®</sup> procedure reference: the UCM procedure. Version 7 & 8. Cary, NC: SAS. 1617-1691 p
- SAS INSTITUTE INC (SAS). 2006b. SAS/ETS<sup>®</sup> procedure reference: the AUTOREG procedure. Version 7 & 8. Cary, NC: SAS. 301-395 p
- SAS INSTITUTE INC (SAS). 2006c. SAS/ETS<sup>®</sup> procedure reference: the X11 procedure. Version 7 & 8. Cary, NC: SAS. 1151-1216 p

- SAS INSTITUTE INC (SAS). 2007a. SAS<sup>®</sup> software for Windows. Version 9.1.3. Cary, NC: SAS.
- SAS INSTITUTE INC (SAS). 2007b. SAS<sup>®</sup> Help and documentation (electronic): goodness-of-fit statistics. Version 9.1.3. Cary, NC: SAS.
- SIN, C. & LAM, K. 2004. The information content of implied volatility: a goodness-of-fit versus an end-user approaches [sic].  
<http://207.36.165.114/NewOrleans/Papers/4402213.pdf> Date of access: 27 Mar. 2006.
- SMITTEN, R. 2005. Trade like Jesse Livermore. Hoboken, NJ: Wiley. 258 p.
- SOUTH AFRICAN FUTURES EXCHANGE (SAFEX). 2006. History of SAFEX.  
<http://www.safex.co.za/ap/recap.asp> Date of access: 22 Dec. 2006.
- SOROS, G. 2003. The alchemy of finance. Hoboken, NJ: Wiley. 391 p.
- SPEARMAN, C. 1904. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72-101, Jan.
- STRONG, R.A. 2002. Derivatives: an introduction. Mason, OH: Thomson/South Western. 444 p.
- SYDNEY FUTURES EXCHANGE (SFE). 2006. SFE Overview.  
<http://www.sfe.com.au> Date of access: 22 Dec. 2006.
- THE TOKYO GRAIN EXCHANGE (TGE). 2006. Exchange history.  
<http://www.tge.or.jp/english/index.shtml> Date of access: 22 Dec. 2006.
- THOMAS, R.L. 2005. Using statistics in economics. London: McGraw-Hill. 611 p.
- TSAY, R.S. 2002. Analysis of financial time series. New York, NY: Wiley. 448 p.
- UNITED STATES DEPARTMENT OF AGRICULTURE (USDA). 2007. Oilseeds: markets and trade. Report 08-07, August.  
<http://ffas.usda.gov/oilseeds/circular/2007/January/oilseedsfull0107.pdf> Date of access: 29 Aug. 2007.
- VAN ZYL, L.C. 2006. Verbal communication with the author. Riebeeckstad. (Recording in possession of the author.)
- VENTER, H. 2007. Verbal communication with the author. Potchefstroom. (Recording in possession of the author.)
- WANG, H-J., ZHANG, R-H., COLE, J. & CHAVEZ, F. 1999. El Niño and the related phenomenon Southern Oscillation (ENSO): The largest signal in interannual climate variation. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20):11071-11072, Sept.
- WATSHAM, T.J. 1998. Futures and options in risk management. 2nd ed. London: Thomson Business Press. 606 p.

WEISSMAN, R.L. 2005. *Mechanical trading systems: pairing trader psychology with technical analysis*. Hoboken, NJ: Wiley. 217 p.

WHISTLER, M. 2004. *Pairs trading: capturing profits and hedging risk with statistical arbitrage strategies*. Hoboken, NJ: Wiley. 279 p.

ZELLER, A. & PALM, F.C. 2004. Time series analysis and the simultaneous equation econometric models. (*In* Zeller, A. & Palm, F.C., *eds*. *The structural econometric time series analysis approach*. New York, NY: Cambridge University Press. p. 3-43.)

# APPENDICES

## STATISTICAL REPORT 1.1

### AUTOCORRELATION TEST ON OLS MODEL BEFORE LAGS (WHITE MAIZE)

#### Durbin-Watson Statistics

Order	DW	Pr < DW	Pr > DW
1	1.1368	<.0001	1.0000
2	1.7393	0.1449	0.8551
3	1.7847	0.2281	0.7719
4	1.9173	0.4958	0.5042
5	2.0394	0.7471	0.2529
6	2.0052	0.7379	0.2621
7	1.8284	0.4376	0.5624
8	1.6791	0.2154	0.7846
9	1.8603	0.5738	0.4262
10	2.1232	0.9397	0.0603
11	1.9105	0.7326	0.2674
12	2.1446	0.7705	0.2295

## STATISTICAL REPORT 1.2

### ESTIMATION OF AUTOREGRESSIVE LAGS (WHITE MAIZE)

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	0.00918	1.000000	*****
1	0.00389	0.423991	*****
2	0.00112	0.122067	**
3	0.000806	0.087805	**
4	0.000154	0.016795	
5	-0.00052	-0.056709	*
6	-0.00041	-0.044437	*
7	0.000326	0.035467	*
8	0.000924	0.100677	**
9	0.000084	0.009193	
10	-0.00124	-0.135107	***
11	-0.00084	-0.091040	**
12	-0.00191	-0.208525	****
13	-0.00159	-0.173037	***

#### Backward Elimination of Autoregressive Terms

Lag	Estimate	t Value	Pr >  t
9	-0.010279	-0.08	0.9330
4	0.018624	0.15	0.8780
6	0.019610	0.17	0.8691
7	-0.016868	-0.16	0.8761
13	-0.036269	-0.33	0.7448
5	0.072395	0.75	0.4577
3	-0.082846	-0.78	0.4359
2	0.078405	0.73	0.4644
11	-0.116201	-1.01	0.3175
8	-0.105047	-1.10	0.2727
10	0.119781	1.26	0.2119
12	0.171345	1.80	0.0753

**Estimates of Autoregressive Parameters**

Lag	Coefficient	Standard Error	t Value
1	-0.423991	0.096000	-4.42

**STATISTICAL REPORT 1.3****ARCH TEST FOR HETEROSKEDASTICITY BEFORE APPLYING AR(1) EGARCH(1,1)-M (WHITE MAIZE)****Q and LM Tests for ARCH Disturbances**

Order	Q	Pr > Q	LM	Pr > LM
1	6.1373	0.0132	5.8831	0.0153
2	6.2185	0.0446	6.6371	0.0362
3	6.7904	0.0789	6.8775	0.0759
4	7.9777	0.0924	7.4978	0.1118
5	10.0331	0.0743	8.5471	0.1286
6	11.0570	0.0866	8.7800	0.1863
7	12.9632	0.0730	10.2649	0.1741
8	14.2118	0.0764	10.7678	0.2152
9	14.2129	0.1149	10.7773	0.2913
10	14.5590	0.1490	10.7872	0.3743
11	15.5061	0.1605	10.7907	0.4610
12	17.1302	0.1448	11.0090	0.5281

**STATISTICAL REPORT 1.4****AUTOCORRELATION TESTS ON THE STANDARDISED RESIDUALS OF THE AR(1) EGARCH(1,1)-M MODELS FITTED FOR EACH OF THE REFERENCE MONTHS (WHITE MAIZE)****Durbin-Watson Statistics**

## January

Order	DW	Pr < DW	Pr > DW
1	2.0707	0.6670	0.3330
2	1.9847	0.5465	0.4535
3	2.0298	0.6653	0.3347
4	1.8179	0.3126	0.6874
5	1.9404	0.5759	0.4241
6	2.0104	0.7465	0.2535
7	1.6863	0.1985	0.8015
8	1.5257	0.0600	0.9400
9	1.6535	0.2058	0.7942
10	2.0658	0.8987	0.1013
11	1.6140	0.2036	0.7964
12	2.0748	0.6489	0.3511

## February

Order	DW	Pr < DW	Pr > DW
1	2.0280	0.5903	0.4097
2	2.0496	0.6651	0.3349
3	2.0141	0.6377	0.3623
4	1.8178	0.3126	0.6874
5	1.9184	0.5343	0.4657
6	2.0097	0.7454	0.2546
7	1.7266	0.2577	0.7423
8	1.5446	0.0723	0.9277
9	1.7715	0.4037	0.5963
10	2.0403	0.8752	0.1248
11	1.7236	0.3865	0.6135
12	2.0389	0.5788	0.4212

## March

Order	DW	Pr < DW	Pr > DW
1	2.2166	0.8706	0.1294
2	2.0656	0.6925	0.3075
3	2.0876	0.7590	0.2410
4	1.8316	0.3364	0.6636
5	1.9435	0.5817	0.4183
6	2.0259	0.7712	0.2288
7	1.7843	0.3556	0.6444
8	1.5598	0.0835	0.9165
9	1.5776	0.1157	0.8843
10	1.9960	0.8258	0.1742
11	1.6473	0.2534	0.7466
12	1.9572	0.4134	0.5866

## April

Order	DW	Pr < DW	Pr > DW
1	2.0193	0.5741	0.4259
2	2.2744	0.9338	0.0662
3	1.9569	0.5317	0.4683
4	1.8534	0.3755	0.6245
5	1.7454	0.2276	0.7724
6	2.0330	0.7820	0.2180
7	1.8511	0.4810	0.5190
8	1.6836	0.2220	0.7780
9	1.5932	0.1316	0.8684
10	1.9799	0.8050	0.1950
11	1.7068	0.3553	0.6447
12	2.1319	0.7503	0.2497

## May

Order	DW	Pr < DW	Pr > DW
1	1.8104	0.2090	0.7910
2	1.7916	0.2100	0.7900
3	1.8593	0.3493	0.6507
4	1.7991	0.2813	0.7187
5	1.9279	0.5523	0.4477
6	1.9235	0.5867	0.4133
7	1.7038	0.2233	0.7767
8	1.6954	0.2396	0.7604
9	1.7996	0.4572	0.5428
10	2.1148	0.9347	0.0653
11	1.6416	0.2444	0.7556
12	2.0534	0.6076	0.3924

## June

Order	DW	Pr < DW	Pr > DW
1	2.1186	0.7454	0.2546
2	2.0516	0.6686	0.3314
3	2.0271	0.6607	0.3393
4	1.8693	0.4049	0.5951
5	1.9183	0.5341	0.4659
6	1.9002	0.5396	0.4604
7	1.7824	0.3522	0.6478
8	1.5703	0.0919	0.9081
9	1.5733	0.1116	0.8884
10	2.0556	0.8898	0.1102
11	1.6845	0.3154	0.6846
12	1.9545	0.4080	0.5920

## July

Order	DW	Pr < DW	Pr > DW
1	2.0393	0.6110	0.3890
2	1.9891	0.5549	0.4451
3	2.0373	0.6782	0.3218
4	1.8256	0.3260	0.6740
5	1.9524	0.5982	0.4018
6	2.0439	0.7981	0.2019
7	1.6791	0.1889	0.8111
8	1.6213	0.1414	0.8586
9	1.6669	0.2252	0.7748
10	2.1410	0.9493	0.0507
11	1.7035	0.3493	0.6507
12	2.0665	0.6331	0.3669

## August

Order	DW	Pr < DW	Pr > DW
1	2.0447	0.6208	0.3792
2	2.0154	0.6038	0.3962
3	2.0607	0.7172	0.2828
4	1.7450	0.1999	0.8001
5	1.9648	0.6210	0.3790
6	2.0657	0.8281	0.1719
7	1.6479	0.1503	0.8497
8	1.6426	0.1664	0.8336
9	1.6602	0.2153	0.7847
10	2.0738	0.9055	0.0945
11	1.7005	0.3439	0.6561
12	1.9571	0.4132	0.5868

## September

Order	DW	Pr < DW	Pr > DW
1	2.0501	0.6307	0.3693
2	2.0060	0.5866	0.4134
3	2.0594	0.7151	0.2849
4	1.7266	0.1757	0.8243
5	1.9764	0.6419	0.3581
6	2.0556	0.8146	0.1854
7	1.6128	0.1135	0.8865
8	1.6574	0.1855	0.8145
9	1.6329	0.1781	0.8219
10	2.0942	0.9210	0.0790
11	1.6556	0.2667	0.7333
12	1.9392	0.3779	0.6221

## October

Order	DW	Pr < DW	Pr > DW
1	2.0072	0.5513	0.4487
2	2.0142	0.6017	0.3983
3	1.9877	0.5895	0.4105
4	1.7962	0.2764	0.7236
5	1.8608	0.4245	0.5755
6	1.9717	0.6791	0.3209
7	1.6795	0.1894	0.8106
8	1.5258	0.0601	0.9399
9	1.7300	0.3277	0.6723
10	1.9872	0.8146	0.1854
11	1.7285	0.3957	0.6043
12	2.0388	0.5786	0.4214

## November

Order	DW	Pr < DW	Pr > DW
1	1.9705	0.4818	0.5182
2	1.9962	0.5682	0.4318
3	2.0575	0.7120	0.2880
4	1.7300	0.1800	0.8200
5	1.9761	0.6415	0.3585
6	2.0175	0.7579	0.2421
7	1.6183	0.1189	0.8811
8	1.6342	0.1563	0.8437
9	1.5934	0.1318	0.8682
10	2.0803	0.9106	0.0894
11	1.6611	0.2756	0.7244
12	1.9101	0.3227	0.6773

## December

Order	DW	Pr < DW	Pr > DW
1	2.0499	0.6303	0.3697
2	2.0670	0.6949	0.3051
3	1.9758	0.5674	0.4326
4	1.7683	0.2331	0.7669
5	1.8712	0.4442	0.5558
6	1.9402	0.6195	0.3805
7	1.6479	0.1504	0.8496
8	1.5277	0.0612	0.9388
9	1.7611	0.3843	0.6157
10	1.9845	0.8111	0.1889
11	1.6634	0.2794	0.7206
12	1.9443	0.3879	0.6121

**STATISTICAL REPORT 1.5**  
**ARCH TESTS FOR HETEROSKEDASTICITY IN THE STANDARDISED RESIDUALS**  
**AFTER APPLYING THE AR(1) EGARCH(1,1)-M MODEL FOR EACH OF THE**  
**REFERENCE MONTHS (WHITE MAIZE)**

**Q and LM Tests for ARCH Disturbances**

**January**

Order	Q	Pr > Q	LM	Pr > LM
1	1.0509	0.3053	1.1011	0.2940
2	1.1598	0.5600	1.2999	0.5221
3	3.1783	0.3649	2.5962	0.4582
4	3.1877	0.5269	2.5976	0.6273
5	3.1892	0.6708	2.6046	0.7607
6	5.3089	0.5049	5.4874	0.4830
7	5.3360	0.6190	5.5741	0.5903
8	5.4084	0.7132	5.8572	0.6632
9	5.7817	0.7616	6.7041	0.6679
10	6.6807	0.7552	7.3978	0.6874
11	6.8611	0.8102	7.6783	0.7418
12	8.6985	0.7284	7.8964	0.7932

**February**

Order	Q	Pr > Q	LM	Pr > LM
1	1.2992	0.2544	1.3388	0.2472
2	1.3790	0.5018	1.5042	0.4714
3	1.9809	0.5764	1.7135	0.6339
4	2.1421	0.7096	1.8068	0.7712
5	2.4357	0.7862	2.1590	0.8267
6	3.3943	0.7580	3.5499	0.7373
7	3.5445	0.8305	3.5594	0.8289
8	3.5509	0.8952	3.5871	0.8923
9	3.8194	0.9229	3.9275	0.9161
10	4.5754	0.9177	4.4309	0.9258
11	4.6166	0.9483	4.4539	0.9547
12	4.6197	0.9695	4.7460	0.9659

**March**

Order	Q	Pr > Q	LM	Pr > LM
1	0.1832	0.6686	0.2177	0.6408
2	0.1866	0.9109	0.2182	0.8967
3	1.4249	0.6997	0.9715	0.8082
4	1.5647	0.8151	1.0093	0.9084
5	1.6402	0.8963	1.2300	0.9420
6	3.4273	0.7536	3.6361	0.7258
7	4.1003	0.7682	3.8720	0.7944
8	4.1107	0.8470	3.8787	0.8679
9	4.2549	0.8938	4.2021	0.8976
10	4.2633	0.9347	4.3580	0.9298
11	4.3832	0.9573	4.4227	0.9559
12	4.3895	0.9753	4.7618	0.9655

**April**

Order	Q	Pr > Q	LM	Pr > LM
1	2.7163	0.0993	2.6997	0.1004
2	3.3688	0.1856	3.8437	0.1463
3	3.9218	0.2700	3.9256	0.2696
4	4.3267	0.3636	4.2176	0.3774
5	4.3544	0.4996	4.2194	0.5183
6	5.3750	0.4967	5.2711	0.5095
7	6.6912	0.4617	8.0429	0.3288
8	7.8811	0.4452	8.1766	0.4164
9	8.7880	0.4571	9.4051	0.4007
10	9.7167	0.4657	9.4710	0.4881
11	9.7339	0.5545	9.5817	0.5684
12	11.485	0.4878	9.7784	0.6354

**May**

Order	Q	Pr > Q	LM	Pr > LM
1	0.6808	0.4093	0.8148	0.3667
2	0.7043	0.7032	0.8885	0.6413
3	0.8688	0.8329	0.8892	0.8280
4	1.6144	0.8062	1.1869	0.8802
5	1.8295	0.8722	1.5492	0.9073
6	3.3945	0.7580	3.7524	0.7101
7	5.2425	0.6304	4.3035	0.7442
8	5.9364	0.6544	5.3006	0.7250
9	6.0063	0.7393	5.3064	0.8068
10	6.9810	0.7272	5.7722	0.8340
11	7.0522	0.7948	6.6452	0.8270
12	8.3117	0.7603	7.1838	0.8452

**June**

Order	Q	Pr > Q	LM	Pr > LM
1	0.0021	0.9634	0.0000	0.9999
2	1.0445	0.5932	1.0416	0.5941
3	2.7674	0.4289	2.0943	0.5531
4	2.7674	0.5975	2.2591	0.6882
5	3.0088	0.6986	2.5039	0.7759
6	4.6029	0.5956	4.8503	0.5631
7	5.3422	0.6183	5.2710	0.6269
8	5.4738	0.7059	5.2719	0.7282
9	5.6148	0.7778	5.7166	0.7679
10	5.6365	0.8448	6.2509	0.7938
11	5.6400	0.8963	6.4746	0.8399
12	5.7506	0.9281	6.6627	0.8791

## July

Order	Q	Pr > Q	LM	Pr > LM
1	1.5652	0.2109	1.6170	0.2035
2	1.9192	0.3830	2.1711	0.3377
3	2.5864	0.4599	2.3523	0.5026
4	2.6774	0.6132	2.3980	0.6630
5	2.9295	0.7109	2.6570	0.7527
6	4.0163	0.6745	4.1133	0.6613
7	4.4060	0.7320	4.1250	0.7653
8	4.8002	0.7787	4.5830	0.8011
9	5.2664	0.8105	5.0421	0.8306
10	7.0906	0.7169	6.3885	0.7816
11	7.1037	0.7906	6.4201	0.8439
12	8.4191	0.7516	6.9140	0.8632

## August

Order	Q	Pr > Q	LM	Pr > LM
1	0.1915	0.6617	0.2193	0.6395
2	1.1951	0.5501	1.2353	0.5392
3	1.5798	0.6640	1.3751	0.7114
4	1.6218	0.8049	1.3753	0.8485
5	1.8121	0.8745	1.5850	0.9031
6	3.3397	0.7652	3.2769	0.7734
7	3.3426	0.8516	3.3381	0.8521
8	3.6942	0.8836	4.0392	0.8536
9	3.7786	0.9254	4.1672	0.9001
10	4.7088	0.9098	4.5809	0.9174
11	4.7260	0.9437	4.6837	0.9455
12	4.8617	0.9624	4.6846	0.9677

## September

Order	Q	Pr > Q	LM	Pr > LM
1	0.3680	0.5441	0.4180	0.5179
2	1.3207	0.5167	1.4359	0.4878
3	1.8511	0.6039	1.6235	0.6541
4	1.8922	0.7556	1.6244	0.8044
5	1.9598	0.8547	1.7030	0.8885
6	3.2467	0.7773	3.1747	0.7866
7	3.3242	0.8535	3.1786	0.8680
8	3.7816	0.8763	4.0102	0.8562
9	3.9990	0.9115	4.2633	0.8932
10	4.5291	0.9203	4.4031	0.9273
11	4.5335	0.9516	4.4061	0.9565
12	4.5568	0.9712	4.7632	0.9654

## October

Order	Q	Pr > Q	LM	Pr > LM
1	1.6768	0.1954	1.6912	0.1934
2	2.0730	0.3547	2.3033	0.3161
3	3.1318	0.3717	2.6382	0.4508
4	3.5242	0.4742	2.9173	0.5718
5	3.8097	0.5771	3.1621	0.6750
6	4.9408	0.5514	4.8498	0.5632
7	5.3609	0.6160	4.8498	0.6783
8	5.3612	0.7184	4.8659	0.7718
9	6.0690	0.7330	5.8464	0.7552
10	6.1234	0.8048	5.8922	0.8242
11	6.1680	0.8619	5.9365	0.8776
12	6.1954	0.9059	6.6382	0.8806

## November

Order	Q	Pr > Q	LM	Pr > LM
1	0.3069	0.5796	0.3362	0.5620
2	1.5834	0.4531	1.6349	0.4416
3	1.9518	0.5825	1.7411	0.6278
4	1.9534	0.7443	1.7664	0.7786
5	2.0796	0.8380	1.9263	0.8593
6	3.6777	0.7202	3.8140	0.7018
7	3.7321	0.8101	3.8248	0.7997
8	4.5901	0.8004	5.3427	0.7204
9	4.7620	0.8545	5.5540	0.7836
10	5.1852	0.8785	5.5757	0.8496
11	5.1948	0.9214	5.5764	0.9001
12	5.1976	0.9511	5.7500	0.9282

## December

Order	Q	Pr > Q	LM	Pr > LM
1	0.6316	0.4268	0.6762	0.4109
2	0.6592	0.7192	0.7326	0.6933
3	1.3349	0.7209	0.9757	0.8071
4	1.5939	0.8099	1.0640	0.8999
5	1.7457	0.8831	1.3751	0.9270
6	4.2625	0.6412	4.7552	0.5756
7	4.5120	0.7193	4.7604	0.6892
8	4.5803	0.8014	4.7649	0.7824
9	4.8700	0.8455	5.2285	0.8140
10	5.0350	0.8888	5.2822	0.8715
11	5.0743	0.9275	5.4976	0.9047
12	5.2191	0.9503	5.7504	0.9281

**STATISTICAL REPORT 1.6**  
**AUTOCORRELATION TEST ON OLS MODEL BEFORE LAGS (YELLOW MAIZE)**

**Durbin-Watson Statistics**

Order	DW	Pr < DW	Pr > DW
1	1.1049	<.0001	1.0000
2	1.7327	0.1378	0.8622
3	1.8176	0.2786	0.7214
4	1.8830	0.4305	0.5695
5	1.8819	0.4646	0.5354
6	2.0413	0.7943	0.2057
7	2.0190	0.7762	0.2238
8	1.6525	0.1790	0.8210
9	1.7448	0.3542	0.6458
10	2.0086	0.8409	0.1591
11	2.0050	0.8595	0.1405
12	2.2629	0.9115	0.0885

**STATISTICAL REPORT 1.7**  
**ESTIMATION OF AUTOREGRESSIVE LAGS (YELLOW MAIZE)**

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	0.00540	1.000000	*****
1	0.00241	0.447484	*****
2	0.000719	0.133265	***
3	0.000447	0.082756	**
4	0.000249	0.046202	*
5	0.000186	0.034425	*
6	-0.00028	-0.051266	*
7	-0.00026	-0.049047	*
8	0.000689	0.127781	***
9	0.000424	0.078508	**
10	-0.00034	-0.063042	*
11	-0.00071	-0.131347	***
12	-0.00141	-0.260832	*****
13	-0.00121	-0.223759	****

**Backward Elimination of Autoregressive Terms**

Lag	Estimate	t Value	Pr >  t
4	-0.007254	-0.06	0.9534
9	0.009387	0.08	0.9393
11	-0.020471	-0.17	0.8653
13	0.040001	0.36	0.7201
5	-0.035177	-0.33	0.7398
7	0.070649	0.61	0.5439
10	0.072622	0.76	0.4521
3	-0.094777	-0.91	0.3665
2	0.074927	0.73	0.4688
6	0.100202	1.08	0.2830
8	-0.158540	-1.72	0.0883

**Estimates of Autoregressive Parameters**

Lag	Coefficient	Standard Error	t Value
1	-0.420479	0.09	-4.49
12	0.205603	0.09	2.20

**STATISTICAL REPORT 1.8****ARCH TEST FOR HETEROSKEDASTICITY BEFORE APPLYING AR(1) EGARCH(1,1)-M (YELLOW MAIZE)****Q and LM Tests for ARCH Disturbances**

Order	Q	Pr > Q	LM	Pr > LM
1	2.3641	0.1242	2.3974	0.1215
2	3.2264	0.1993	3.5984	0.1654
3	3.8917	0.2734	3.7840	0.2858
4	3.9206	0.4169	3.7846	0.4359
5	5.3074	0.3795	5.0990	0.4039
6	6.2348	0.3974	5.2995	0.5060
7	7.1022	0.4183	5.8481	0.5576
8	7.6194	0.4715	6.1137	0.6345
9	9.3657	0.4042	7.2776	0.6082
10	10.0580	0.4354	7.5683	0.6709
11	10.0580	0.5252	7.7569	0.7349
12	10.8623	0.5408	7.8708	0.7951

**STATISTICAL REPORT 1.9****AUTOCORRELATION TESTS ON THE STANDARDISED RESIDUALS OF THE AR(1) EGARCH(1,1)-M MODELS FITTED FOR EACH OF THE REFERENCE MONTHS (YELLOW MAIZE)****Durbin-Watson Statistics**

January				February				March			
Order	DW	Pr < DW	Pr > DW	Order	DW	Pr < DW	Pr > DW	Order	DW	Pr < DW	Pr > DW
1	1.9521	0.4470	0.5530	1	2.0674	0.6612	0.3388	1	2.0282	0.5907	0.4093
2	2.0976	0.7440	0.2560	2	2.0068	0.5880	0.4120	2	2.0747	0.7076	0.2924
3	2.0392	0.6816	0.3184	3	1.9579	0.5336	0.4664	3	2.0437	0.6892	0.3108
4	1.7998	0.2823	0.7177	4	1.9251	0.5107	0.4893	4	1.6897	0.1329	0.8671
5	1.7249	0.1988	0.8012	5	1.7945	0.3052	0.6948	5	1.5864	0.0637	0.9363
6	2.1420	0.9095	0.0905	6	2.1689	0.9301	0.0699	6	2.1802	0.9376	0.0624
7	2.0811	0.8550	0.1450	7	2.0182	0.7751	0.2249	7	2.1095	0.8839	0.1161
8	1.6556	0.1831	0.8169	8	1.7234	0.2843	0.7157	8	1.8122	0.4453	0.5547
9	1.6549	0.2079	0.7921	9	1.6686	0.2277	0.7723	9	1.7148	0.3014	0.6986
10	1.8556	0.6021	0.3979	10	1.9645	0.7838	0.2162	10	1.8901	0.6649	0.3351
11	1.8344	0.5995	0.4005	11	1.7548	0.4461	0.5539	11	1.8867	0.6934	0.3066
12	2.1921	0.8377	0.1623	12	2.1635	0.7988	0.2012	12	2.1896	0.8345	0.1655

## April

Order	DW	Pr < DW	Pr > DW
1	1.9535	0.4497	0.5503
2	2.0919	0.7352	0.2648
3	2.0705	0.7329	0.2671
4	1.8168	0.3108	0.6892
5	1.7292	0.2045	0.7955
6	2.1166	0.8864	0.1136
7	2.0932	0.8678	0.1322
8	1.6526	0.1792	0.8208
9	1.6396	0.1869	0.8131
10	1.8882	0.6616	0.3384
11	1.8250	0.5816	0.4184
12	2.1760	0.8164	0.1836

## May

Order	DW	Pr < DW	Pr > DW
1	1.9733	0.4870	0.5130
2	2.0471	0.6608	0.3392
3	2.0356	0.6753	0.3247
4	1.8360	0.3442	0.6558
5	1.7003	0.1671	0.8329
6	2.1488	0.9152	0.0848
7	2.0807	0.8545	0.1455
8	1.6750	0.2096	0.7904
9	1.6897	0.2600	0.7400
10	1.8310	0.5555	0.4445
11	1.8984	0.7129	0.2871
12	2.1029	0.7007	0.2993

## June

Order	DW	Pr < DW	Pr > DW
1	1.9618	0.4652	0.5348
2	2.1654	0.8365	0.1635
3	2.0235	0.6544	0.3456
4	1.8233	0.3219	0.6781
5	1.6871	0.1515	0.8485
6	2.1640	0.9266	0.0734
7	2.1235	0.8965	0.1035
8	1.6750	0.2096	0.7904
9	1.6340	0.1796	0.8204
10	1.8662	0.6218	0.3782
11	1.8244	0.5804	0.4196
12	2.1915	0.8368	0.1632

## July

Order	DW	Pr < DW	Pr > DW
1	1.9102	0.3696	0.6304
2	2.0585	0.6805	0.3195
3	1.9635	0.5442	0.4558
4	1.8207	0.3174	0.6826
5	1.8459	0.3966	0.6034
6	2.1187	0.8885	0.1115
7	2.1430	0.9124	0.0876
8	1.6337	0.1557	0.8443
9	1.6745	0.2364	0.7636
10	1.9126	0.7035	0.2965
11	1.7287	0.3961	0.6039
12	2.1353	0.7558	0.2442

## August

Order	DW	Pr < DW	Pr > DW
1	1.9973	0.5327	0.4673
2	2.0883	0.7295	0.2705
3	2.0610	0.7177	0.2823
4	1.8044	0.2900	0.7100
5	1.7155	0.1863	0.8137
6	2.1546	0.9197	0.0803
7	2.1042	0.8788	0.1212
8	1.6723	0.2057	0.7943
9	1.6345	0.1802	0.8198
10	1.8750	0.6380	0.3620
11	1.8372	0.6047	0.3953
12	2.1663	0.8029	0.1971

## September

Order	DW	Pr < DW	Pr > DW
1	1.9910	0.5208	0.4792
2	2.1093	0.7616	0.2384
3	2.0340	0.6727	0.3273
4	1.7581	0.2182	0.7818
5	1.7006	0.1674	0.8326
6	2.1246	0.8941	0.1059
7	2.0949	0.8696	0.1304
8	1.6525	0.1790	0.8210
9	1.6717	0.2322	0.7678
10	1.8742	0.6365	0.3635
11	1.8325	0.5958	0.4042
12	2.1785	0.8198	0.1802

## October

Order	DW	Pr < DW	Pr > DW
1	1.9428	0.4295	0.5705
2	2.0932	0.7372	0.2628
3	2.0424	0.6870	0.3130
4	1.8415	0.3541	0.6459
5	1.7292	0.2046	0.7954
6	2.1211	0.8908	0.1092
7	2.0731	0.8460	0.1540
8	1.6489	0.1744	0.8256
9	1.6497	0.2005	0.7995
10	1.8639	0.6175	0.3825
11	1.8423	0.6141	0.3859
12	2.1876	0.8319	0.1681

## November

Order	DW	Pr < DW	Pr > DW
1	1.8601	0.2836	0.7164
2	1.9007	0.3882	0.6118
3	1.9565	0.5308	0.4692
4	1.8501	0.3697	0.6303
5	1.7542	0.2406	0.7594
6	2.0754	0.8404	0.1596
7	2.0865	0.8608	0.1392
8	1.7163	0.2726	0.7274
9	1.7887	0.4363	0.5637
10	1.9450	0.7552	0.2448
11	1.8774	0.6774	0.3226
12	2.1374	0.7591	0.2409

## December

Order	DW	Pr < DW	Pr > DW
1	1.8764	0.3105	0.6895
2	1.9797	0.5370	0.4630
3	1.9494	0.5173	0.4827
4	1.7835	0.2563	0.7437
5	1.6784	0.1418	0.8582
6	2.0923	0.8605	0.1395
7	2.1294	0.9015	0.0985
8	1.7128	0.2670	0.7330
9	1.7412	0.3478	0.6522
10	1.8820	0.6505	0.3495
11	1.9172	0.7431	0.2569
12	2.1458	0.7724	0.2276

**STATISTICAL REPORT 1.10**  
**ARCH TESTS FOR HETEROSKEDASTICITY IN THE STANDARDISED RESIDUALS**  
**AFTER APPLYING THE AR(1) EGARCH(1,1)-M MODEL FOR EACH OF THE**  
**REFERENCE MONTHS (YELLOW MAIZE)**

**Q and LM Tests for ARCH Disturbances**

**January**

Order	Q	Pr > Q	LM	Pr > LM
1	1.6633	0.1972	1.4335	0.2312
2	3.2845	0.1935	3.0336	0.2194
3	4.6729	0.1974	3.9770	0.2640
4	4.8187	0.3064	4.3596	0.3595
5	4.8188	0.4384	4.6278	0.4630
6	5.9633	0.4273	5.1706	0.5221
7	6.2698	0.5086	5.8196	0.5610
8	6.4092	0.6015	5.8198	0.6674
9	6.8760	0.6500	6.5999	0.6787
10	7.8207	0.6463	6.7584	0.7480
11	11.1184	0.4334	9.9427	0.5356
12	12.8965	0.3766	10.0022	0.6158

**February**

Order	Q	Pr > Q	LM	Pr > LM
1	0.6702	0.4130	0.5476	0.4593
2	3.9135	0.1413	3.4674	0.1766
3	6.3432	0.0961	5.4273	0.1431
4	7.2367	0.1239	6.4590	0.1674
5	8.6072	0.1258	6.5719	0.2545
6	9.4338	0.1506	6.9773	0.3230
7	9.5888	0.2131	6.9942	0.4295
8	9.5967	0.2945	6.9954	0.5371
9	9.6163	0.3824	7.3651	0.5992
10	10.6281	0.3872	8.1310	0.6160
11	11.7727	0.3810	9.7069	0.5569
12	13.3091	0.3470	9.9529	0.6201

**March**

Order	Q	Pr > Q	LM	Pr > LM
1	1.3985	0.2370	1.1743	0.2785
2	3.1610	0.2059	2.7618	0.2513
3	5.9018	0.1165	4.8115	0.1861
4	6.2907	0.1785	4.9689	0.2905
5	6.4624	0.2638	5.5558	0.3519
6	6.4637	0.3733	5.5623	0.4740
7	8.9188	0.2585	6.8839	0.4411
8	9.9878	0.2659	7.4789	0.4859
9	10.9838	0.2768	8.7539	0.4603
10	14.8680	0.1369	9.9645	0.4436
11	15.8340	0.1474	11.6072	0.3939
12	16.4743	0.1705	11.6379	0.4752

**April**

Order	Q	Pr > Q	LM	Pr > LM
1	1.5546	0.2125	1.3434	0.2464
2	4.2780	0.1178	4.0210	0.1339
3	6.3886	0.0942	5.3789	0.1461
4	6.5426	0.1621	5.7172	0.2213
5	6.5690	0.2547	6.4018	0.2691
6	8.8447	0.1825	7.7290	0.2586
7	9.0753	0.2473	8.3641	0.3016
8	9.1592	0.3290	8.4907	0.3871
9	9.8217	0.3651	9.5974	0.3840
10	11.2567	0.3379	9.9673	0.4434
11	13.8541	0.2412	11.9739	0.3656
12	15.5745	0.2115	11.9772	0.4475

**May**

Order	Q	Pr > Q	LM	Pr > LM
1	2.3246	0.1273	2.0577	0.1514
2	6.1085	0.0472	6.0649	0.0482
3	8.0475	0.0450	6.8798	0.0758
4	8.7372	0.0680	7.7495	0.1012
5	9.2284	0.1003	10.0770	0.0731
6	10.8081	0.0945	10.3391	0.1111
7	11.4599	0.1198	11.4900	0.1186
8	11.5689	0.1715	12.0721	0.1480
9	12.6623	0.1785	12.5168	0.1857
10	14.7544	0.1413	13.3029	0.2072
11	16.3411	0.1289	14.1984	0.2222
12	16.9709	0.1507	14.2656	0.2841

**June**

Order	Q	Pr > Q	LM	Pr > LM
1	1.8264	0.1766	1.5808	0.2086
2	3.7007	0.1572	3.4526	0.1779
3	5.7502	0.1244	4.8450	0.1835
4	5.8220	0.2128	4.8451	0.3036
5	5.8994	0.3161	5.3608	0.3735
6	7.0277	0.3183	6.1390	0.4078
7	7.0891	0.4197	6.1970	0.5169
8	7.1752	0.5179	6.1973	0.6251
9	8.9943	0.4378	9.1582	0.4228
10	11.6016	0.3126	10.3555	0.4099
11	13.5225	0.2606	11.9813	0.3651
12	15.0226	0.2402	11.9893	0.4465

## July

Order	Q	Pr > Q	LM	Pr > LM
1	1.6288	0.2019	1.4099	0.2351
2	2.4041	0.3006	2.2058	0.3319
3	3.6202	0.3055	3.1896	0.3633
4	3.8290	0.4296	3.7831	0.4362
5	3.8723	0.5679	4.2587	0.5128
6	5.6517	0.4633	5.1994	0.5185
7	5.6532	0.5808	5.2442	0.6302
8	5.7850	0.6713	5.5357	0.6991
9	5.9921	0.7407	6.1006	0.7298
10	6.3414	0.7858	6.1553	0.8021
11	8.8326	0.6373	7.9503	0.7177
12	10.3604	0.5844	8.1968	0.7696

## August

Order	Q	Pr > Q	LM	Pr > LM
1	2.4725	0.1159	2.1822	0.1396
2	4.4635	0.1073	4.4035	0.1106
3	7.2273	0.0650	6.2196	0.1014
4	7.2278	0.1243	6.3688	0.1732
5	7.2286	0.2042	6.7198	0.2423
6	8.2703	0.2190	7.4222	0.2836
7	8.5340	0.2879	8.0070	0.3320
8	8.5504	0.3816	8.0962	0.4241
9	9.4278	0.3988	9.4281	0.3987
10	10.7234	0.3795	9.6862	0.4684
11	13.9728	0.2345	12.6347	0.3179
12	16.4873	0.1699	12.7005	0.3912

## September

Order	Q	Pr > Q	LM	Pr > LM
1	1.3795	0.2402	1.1776	0.2778
2	3.1008	0.2122	2.8157	0.2447
3	5.2907	0.1517	4.5056	0.2118
4	5.3228	0.2558	4.5171	0.3405
5	5.4255	0.3662	5.1576	0.3970
6	7.5106	0.2762	6.7008	0.3494
7	7.7053	0.3593	6.9086	0.4385
8	7.7320	0.4601	6.9575	0.5412
9	8.7270	0.4628	9.2169	0.4175
10	10.1275	0.4294	9.7226	0.4652
11	13.9799	0.2341	12.4305	0.3322
12	16.2597	0.1796	12.4834	0.4077

## October

Order	Q	Pr > Q	LM	Pr > LM
1	1.2852	0.2569	1.0958	0.2952
2	4.0545	0.1317	3.7286	0.1550
3	5.7223	0.1259	4.8283	0.1848
4	5.9353	0.2040	5.1761	0.2697
5	6.0737	0.2991	6.1684	0.2902
6	8.8394	0.1828	7.6694	0.2633
7	9.0733	0.2474	8.2330	0.3125
8	9.2381	0.3226	8.3347	0.4015
9	9.6859	0.3765	9.0944	0.4286
10	11.2205	0.3406	9.6102	0.4753
11	13.5459	0.2591	11.3844	0.4116
12	15.2907	0.2259	11.4079	0.4943

## November

Order	Q	Pr > Q	LM	Pr > LM
1	0.1422	0.7061	0.0890	0.7654
2	3.2186	0.2000	2.6343	0.2679
3	4.7711	0.1893	4.1185	0.2490
4	7.2613	0.1227	6.2997	0.1779
5	7.3373	0.1967	7.0950	0.2137
6	10.5823	0.1022	8.5542	0.2002
7	10.9327	0.1416	9.2639	0.2343
8	11.0435	0.1992	9.4055	0.3092
9	11.0678	0.2711	9.5453	0.3885
10	12.7172	0.2399	9.7677	0.4611
11	14.8901	0.1876	10.8598	0.4551
12	16.4776	0.1703	11.3052	0.5030

## December

Order	Q	Pr > Q	LM	Pr > LM
1	1.3936	0.2378	1.2013	0.2731
2	5.6375	0.0597	5.3006	0.0706
3	7.6887	0.0529	6.4264	0.0926
4	8.2185	0.0839	6.9716	0.1374
5	9.2654	0.0989	10.1885	0.0701
6	12.0368	0.0612	11.2030	0.0823
7	13.0968	0.0698	12.4677	0.0862
8	13.3550	0.1002	13.0247	0.1110
9	14.7945	0.0967	14.0294	0.1213
10	17.3719	0.0665	14.9810	0.1328
11	19.0185	0.0608	15.3614	0.1665
12	20.2222	0.0630	15.3617	0.2223

## INDEX

## A

- (Australian Bureau Meteorology) **ABM**, 21, 22, 225  
 (Augmented Dicky-Fuller) **ADF**, 141, 142, 143, 145, 146  
 (Akaike) **AIC**, 87, See also **SBC**  
 (Agricultural Markets Division) **AMD**, 30  
**ARCH**, 44, 59, 60, 61, 228, 239, 241, 244, 246, See also **GARCH**  
**ARCHTEST**, 76, 79, 88  
 (Autoregressive Integrated Moving Average) **ARIMA**, 6, 7, 52, 54, 56, 58, 59, 69, 71, 72, 73, 74, 92, 93, 94, 95, 96, 97, 99, 100, 102, 105, 108, 111, 112, 122, 133, 217, 219, 227, See also **ARMA** and **SARIMA**  
 (Autoregressive Moving Average) **ARMA**, 7, 54, 56, 57, 58, 59, 69, 217, 219 See also **ARIMA** and **SARIMA**
- Autocorrelation**, 58, 76, 77, 78, 80, 82, 85, 86, 87, 88, 89, 90, 100  
**AUTOREG**, 76, 79, 87, 88, 235  
**Autoregression**, 79  
 Autoregressive, 57, 60, 66, 76, 78, 79, 82, 87, 88, 138, 147, 232

## B

- Beta**, 6, 8, 71, 135, 137, 138, 144, 157, 164, 165, 166, 167, 168, 169, 188, 220  
**Bollinger bands**, 41, 43, 44, 46, 47, 48, 49, 51, 52, 197, 200, 201, 209, 210, 211, 214, 215, 216, 218, 222, 223, See also **Technical analysis**  
**Brownian motion**, 54, 233

## C

- (Capital asset pricing model) **CAPM**, 62, 81  
 (Chicago Board of Trade) **CBOT**, 26, 27, 224, 226  
 (Chicago Mercantile Exchange) **CME**, 27, 226  
 (Chande Momentum Oscillator) **CMO**, 36, 37, 197, See also **Technical analysis**  
**Co-integration**, 5, 6, 8, 135, 137, 138, 140, 141, 142, 144, 145, 146, 147, 148, 149, 151, 152, 157, 161, 164, 166, 169, 187, 220, 224, 225, 230, 231, 232

- Contango**, 23, 24  
**Covariance**, 64, 154, 166

## D

- (Durbin Watson) **DW**, 145, 238, 239, 240, 243, 245, 246

## E

- (Exponential Generalised Autoregressive Conditional Heteroskedasticity) **EGARCH**, 62, 79, 80, 81, 82, 83, 88, 89, 90, 233, 239, 241, 244, 245, 246, See also **GARCH**  
**Eigenvalue**, 147  
**El Niño**, 21, 22, 229, 236, See also **ENSO**  
 (El Niño-Southern Oscillation) **ENSO**, 21, 22, 236  
**Ergodic**, 64  
**Ergodicity**, 64  
**EViews**, 93, 95, 141, 142, 148, 149, 156, 234  
 (Exponentially Weighted Moving Average) **EWMA**, 41, 43, 44, See also **SMA** and **MA**

## G

- (Generalised Autoregressive Conditional Heteroskedasticity) **GARCH**, 6, 7, 41, 44, 52, 54, 55, 56, 57, 59, 60, 61, 62, 63, 64, 65, 69, 71, 72, 73, 74, 77, 78, 80, 81, 85, 86, 92, 99, 115, 133, 217, 219, 228, 233, See also **EGARCH**  
**Granger**, 138, 139, 140, 141, 145

## H

- Henderson-curve**, 93, 94, 95, 99, 100, 102, 103, 104, 106, 107, 108, 110, 114, 115, 133, 136, 170, 171, 176, 181, 219  
**Heteroskedasticity**, 44, 55, 59, 76, 77, 79, 80, 81, 83, 85, 87, 88, 90, 225, 227, 232, 233  
**Homoskedastic**, 79

## J

- Jarque-Bera**, 83, 90  
**Johansen**, 145, 146, 147, 148, 149, 151, 224  
 (Johannesburg Stock Exchange) **JSE**, 30, 231

**K**

**Kurtosis**, 60, 77, 83, 86, 90. See also **Skewness**

**L**

**La Niña**, 22  
**(London Commodity Exchange) LCE**, 28  
**(Lagrange multiplier) LM**, 79, 83, 88, 239, 241, 242, 244, 246, 247, 248

**M**

**(Moving Average) MA**, 44, 45, 46, 57, 58, 100, 197, 200, 212, 213, 214, 216, 218, 222, 223, 227, 230, 231. See also **Technical analysis**, **EWMA** and **SMA**  
**(Moving Average Convergence/Divergence) MACD**, 41, 49, 50, 51, 52, 197, 200, 201, 211, 212, 213, 214, 215, 216, 218, 222. See also **Technical analysis**  
**Maize**, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 26, 27, 28, 29, 30, 72, 73, 74, 75, 76, 77, 78, 79, 80, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 99, 100, 102, 103, 104, 105, 106, 107, 108, 110, 111, 112, 113, 114, 115, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 131, 132, 133, 135, 138, 149, 151, 155, 165, 167, 168, 169, 170, 171, 176, 181, 182, 186, 187, 188, 189, 190, 191, 197, 199, 200, 202, 204, 219, 220, 221, 228, 232, 235  
**Multivariate**, 62, 226

**N**

**(New York Board of Trade) NYBOT**, 28, 233  
**(New York Cotton Exchange) NYCE**, 28

**O**

**(Ordinary least square) OLS**, 76, 78, 87, 156, 238, 243  
**Oscillators**, 34  
**Overbought**, 33, 34, 35, 36, 40, 47, 49, 51, 70, 200, 215. See also **Technical analysis**  
**Oversold**, 33, 34, 35, 36, 40, 47, 49, 51, 70, 200, 215. See also **Technical analysis**

**P**

**Pairs**, 1, 2, 3, 4, 5, 6, 7, 8, 10, 25, 26, 27, 30, 31, 32, 40, 41, 44, 51, 53, 59, 69, 70, 71, 72, 74, 85, 92, 93, 94, 102, 103, 106, 108, 110, 115, 121, 126, 132, 133, 134, 135, 136, 137, 138, 140, 141, 144, 152, 155, 156, 157, 158, 159, 161, 162, 163, 164, 166, 167, 168, 169, 170, 171, 175, 176, 181, 182, 183, 185, 186, 187, 188, 189, 190, 191, 196, 197, 198, 199, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 220, 221, 222, 223, 224, 227, 232  
**Pairs trading**, 1, 2, 3, 5, 54, 190, 191, 198, 211, 218, 223, 224  
**Pearson's correlation**, 8, 135, 137, 138, 144, 152, 154, 155, 156, 157, 159, 160, 161, 165, 220

**R**

**(Relative Strength Index) RSI**, 32, 34, 35, 36, 37, 39, 40, 47, 52, 197, 200, 206, 207, 208, 214, 216, 218, 222, 223. See also **Technical analysis**

**S**

**(South African Futures Exchange) SAFEX**, 6, 9, 11, 13, 14, 15, 16, 17, 18, 26, 30, 53, 72, 73, 149, 187, 202, 205, 207, 209, 210, 214, 218, 224, 228, 231, 235  
**(South African Grain Information Service) SAGIS**, 12, 15, 235  
**(Seasonal Autoregressive Integrated Moving Average) SARIMA**, 93. See also **ARIMA** and **ARMA**  
**SAS**, 56, 66, 67, 68, 69, 74, 76, 79, 82, 87, 88, 89, 92, 93, 94, 95, 96, 97, 98, 100, 101, 105, 116, 117, 119, 123, 124, 128, 129, 156, 235  
**(Schwarz's Bayesian information criterion) SBC**, 87, 88. See also **AIC**  
**Seasonal patterns**, 7, 11, 31, 41, 52, 53, 56, 65, 69, 70, 71, 72, 77, 85, 92, 93, 95, 99, 110, 114, 115, 118, 120, 121, 122, 126, 129, 131, 132, 133, 134, 136, 137, 140, 169, 170, 176, 180, 182, 186, 188, 196, 197, 219, 220, 222, 223  
**(Sydney Futures Exchange) SFE**, 29, 235  
**Skewness**, 77, 80, 83, 86, 90. See also **Kurtosis**  
**(Simple Moving Average) SMA**, 41, 42, 43, 44, 46, 47. See also **EWMA** and **MA**

**(Southern Oscillation Index) SOI**, 21, 22,

See also **ENSO**, **El Niño** and **La Niña**

**Soybean('s)**, 17, 18, 20, 27, 29, 72, 230

**Spearman's rank correlation**, 8, 71, 135,  
137, 138, 144, 152, 159, 160, 161, 162,  
163, 165, 168, 169, 187, 220, 221

**(Structural Time Series Analysis) STA**, 56,  
59, 65, 69, See also **UCM**

**(Structural Time Series Analyser, Modeller  
and Predictor) STAMP**, 116, 234, See  
also **UCM**

**Stochastic Oscillator**, 197, 207

**Sunflower**, 15, 16, 234

## T

**Technical analysis**, 6, 7, 11, 30, 31, 32, 52,  
71, 110, 189, 196, 197, 203, 204, 206,  
209, 210, 211, 212, 214, 215, 216, 218,  
219, 222, 223, 224, 227, 234, 235, 237

**(Tokyo Grain Exchange) TGE**, 29, 235

**(Time series Regression with ARIMA  
noise, Missing values, and Outliers)  
TRAMO**, 95, 112, 229

## U

**(Unobserved Component Modelling)**

**UCM**, 6, 7, 8, 9, 52, 54, 56, 57, 69, 72, 73,  
74, 115, 116, 118, 122, 126, 127, 131,  
132, 133, 181, 182, 183, 186, 217, 219,  
220, 223, 235, See also **STA** and **STAMP**

**(United States Department of Agriculture)  
USDA**, 16, 17, 236

## W

**Wheat**, 9, 12, 14, 15, 16, 17, 18, 27, 28, 29,  
30, 72, 187

## LIST OF ACRONYMS

ABM	Australian Bureau Meteorology
ADF	Augmented Dicky-Fuller
AIC	Akaike Information criterion
AMD	Agricultural Markets Division
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CAPM	Capital Asset Pricing Model
CBOT	Chicago Board of Trade
CME	Chicago Mercantile Exchange
CMO	Chande Momentum Oscillator
DW	Durbin-Watson
EGARCH	Exponential Generalised Autoregressive Conditional Heteroskedasticity
ENSO	El Niño-Southern Oscillation
EWMA	Exponentially Weighted Moving Average
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
JSE	Johannesburg Stock Exchange
LCE	London Commodity Exchange
LTOM	London Traded Options Market
MA	Moving Average
MACD	Moving Average Convergence/Divergence
MAPE	Mean Absolute Percent Error
NCEC	National Crop Estimate Committee
NYBOT	New York Board of Trade
NYCE	New York Cotton Exchange
OLS	Ordinary Least Squares
QMS	Quantitative Micro Software
RMB	Rand Merchant Bank
RSI	Relative Strength Index
SAFEX	South African Futures Exchange
SAGIS	South African Grain Information Service
SARIMA	Seasonal Autoregressive Integrated Moving Average
SBC	Schwarz's Bayesian Information Criterion
SFE	Sydney Futures Exchange
SMA	Simple Moving Average
SOI	Southern Oscillation Index
STA	Structural Time Series Analysis
STAMP	Structural Time Series Analyser, Modeller and Predictor
TGE	Tokyo Grain Exchange
TRAMO	Time Series Regression with ARIMA noise, Missing values, and Outliers
UCM	Unobserved Component Modelling
USDA	United States Department of Agriculture