行政院國家科學委員會補助專題研究計畫 ■成果報告 □期中進度報告

智慧型金融資訊服務決策支援系統：理財規劃服務多階段動態調適模式之建構

計畫類別：■個別型計畫　□整合型計畫
計畫編號：NSC 97-2410-H-009-037-MY2
執行期間：97 年 08 月 01 日至 99 年 07 月 31 日

執行機構及系所：交通大學資管所

計畫主持人：陳安斌
共同主持人：
計畫參與人員：博士班研究生-兼任助理：許育嘉、蔡文智、楊博文
碩士班研究生-兼任助理：謝璁賦、王怡涵、林益民、葉佩昀、查欣瑜

成果報告類型(依經費核定清單規定繳交)：□精簡報告　■完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：
□赴國外出差或研習心得報告
□赴大陸地區出差或研習心得報告
□出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年■二年後可公開查詢

中　華　民　國　99 年 10 月 29 日

目錄

摘要

以家庭及人生生涯規劃為基礎之理財規劃顧問服務，是近年來逐漸興起之新興金融服務。然而在實務上，家庭及人生的理財規劃需同時考量各種不同的外在因素，如法律、節稅等限制，並且也需配合不同之投資行為與風險偏好選擇不同的投資策略與資產配置。因此一個完美的理財規劃模型，是一個非常複雜的多目標決策問題。傳統上這類問題大多採用線性規劃的方式，來求得最佳解，然而這類的傳統方法學，並無法即時因應外界多變的環境來進行動態的調適，而原本根據某個時間點所做的規劃，也無法在長期時間的考驗下仍然保持最佳化，因此本計畫將利用人工智慧的方法學，建構一個能隨著環境的變動進行動態調適的理財規劃模型。

在第一年，發展出展出以自組織類神經網路為基礎之最佳化動態避險比例估算模型。第二年針對套利策略進行動態知識的發現。最廣泛被使用的套利策略源自於持有成本模型(Cost of Carry)或經濟學的方法。然而實務上，這些方法在處理日內交易、分線交易和跨市場套利有其困難。在本研究中，我們提出基於分類元（XCS）的計算智慧。首先，為了減少資料量，將一分鐘的原始交易資料透過條件過濾成價差關係，接著採用XCS 知識規則的發現去分析特定領域的知識：在到期日指數期貨的價格將會接近現貨價貨，有四個重要的相關因素：基差、價差、到期日期、盤中交易時間，這四個因素被認為是 XCS 建立跨市場套利模式的條件。本研判選擇在台灣期貨交易所(TAIFEX)交易的台股期貨（TX）和在新加坡交易所(SGX)的摩台指(MSCI)作為跨市場研究的標的來實證研究驗證的準確性和盈利能力的模型。

關鍵字：交易規則、高頻分析、日內交易、分類元系統

Abstract

The family or life financial planning is a emerging financial service in recent years. However, it is not easy to make plan that many factors must be considered, such as tax, law and various risk preference. In practice, the financial planning is a complex multi-objective decision problem with many restricted condition. Traditionally, linear program and numerical optimization skill were used to solve these problems. But the investment environment is dynamic and change rapid, that the financial planning should be dynamic adaptive with time. We propose a model which is capable with dynamic adaptive and suitable for long time financial planning.

The result of this project in the first year is to develop a optimal dynamic hedge ratio estimation model using self-organizing map (SOM). In the Second year is to develop a dynamic arbitrage strategy model. The most popular arbitrage strategy is derived from the cost of carry model or by using the econometrics approach. However, these approaches have difficulty in dealing with intra-day 1-minute trading data and capturing inter-market arbitrage opportunity in the real world. In this research, we propose computational intelligence approaches based on the extended classifier system (XCS). First, in order to reduce the amount of data, the original data streams of intra-day 1-minute trading data are filtered by the conditions of variant price spread relation. XCS is then adopted for knowledge rule discovery. After analyzing the property with domain-specific knowledge that the price of index futures will get close to that of spot products at the time the futures mature, four important factors related to bias, price spread, expiry date, and intraday trading timing are considered as the conditions of XCS to build the inter-market arbitrage model. The inter-market spread of the Taiwan Stock Index Futures (TX) traded at the Taiwan Futures Exchange (TAIFEX) and the Morgan Stanley Capital International (MSCI) Taiwan Index Futures traded at the Singapore Exchange Limited (SGX) are chosen for an empirical study to verify the accuracy and profitability of the model.

Keywords: trading rule; high frequency data; intra-day trading; XCS.

1. Introduction

In futures and options markets, if market imperfection or market inefficiency exists, the phenomenon of mispricing can easily occur, creating a price difference between commodities or underlying products (price spread), thereby often leading to the rise of arbitrage opportunities. This phenomenon is more common amongst emerging markets and markets experiencing thinner trading volumes than in mature markets and markets with higher transaction volumes (Wang & Hsu, 2006). Depending on different exchange commodities and exchange markets, trading in the price spread can generally be divided into two types (Moles & Terry, 1997): (1) inter-market (or inter-commodity) spread: two highly related financial products that are traded within the same stock exchange, e.g. arbitrage between stocks and index futures; or two financial products offered in different exchanges covering the same underlying commodity or similar commodities. (2) intra-market (or intra-commodity) spread: arbitrage between products with the same underlying commodity but with different expiry months, e.g. futures contracts for the same index which mature on different months. In general, arbitrage opportunities are rare and difficult to discover, because the calculations are too complex, especially with inter-market arbitrage trading, where expiry dates for different futures contracts, the immediate foreign exchange calculation, and immediate calculation of fair price must all be considered at the same time.

Regardless of the type of arbitrage, when evaluating an opportunity in arbitrage, establishing the fair price of the product and then assessing the magnitude of the price spread are the most important research issues worthy of attention. According to previous studies, the methodologies for detecting arbitrage opportunities can be classified into three categories: the cost of carry model, econometric and behavioral finance, and computational Intelligence approach.

The cost of carry model is the most basic theorem when considering futures arbitrage. However, the actual prices in the index futures markets are generally found to be lower than the theoretical prices predicted by the cost of carry model (Cornell & French, 1983; Figlewski, 1984; Modest & Sundaresan, 1983). This makes the model imperfect in explaining and forecasting price movements in stocks and index futures (Klemkosky & Lee, 1991).

The econometric model considering the arbitrageur behavior can yield a more accurate evaluation of the probability of profiting through arbitrage in practice. Some researchers investigated the inter-market spread trading based on econometrics, such as the spread of West Texas Intermediate (WTI) and Brent Crude (Brent) spread (Dunis et al., 2006; Dunis et al., 2008), the price spread between the Singapore Exchange Limited (SGX) Morgan Stanley Capital International (MSCI) Taiwan Index and the Taiwan Futures Exchange (TAIFEX) Taiwan Stock Index Futures (TX), the price spread between the TAIFEX Taiwan Stock Exchange Electronic Sector Index (TE) and the Financial Sector Index (TF) Futures (Luo, 2002), and the FTSE 100 and the FTSE Mid 250 contract traded on the London International Financial Futures and Options Exchange (LIFFE) (Butterworth & Holmes, 2002). However, most econometric models use time series as a starting point, using only the daily closing prices, and neglects other determining factors and conditions (e.g. time to maturity). Hence, there are some limitations in their capacity to evaluate the probability of arbitrage, especially for intra-day trading data and it is difficult to apply these models to develop a trading decision support system that is capable of high frequency data processing.

In processing high-frequency data for intra-day trading, computational intelligence is a new approach. The real-time updated transaction data are typical data streams that can be processed using temporal data mining skill. A statistical arbitrage trading system for the S&P 500 Futures Index is proposed based on the flexible least squares (FLS), showing that the FLS can be employed as a building block of an algorithmic trading system (Montana et al., 2009).

More recently, many trading decision support systems have been developed based on computational intelligence techniques. There are two main approaches to developing trading decision support system. One is the pricing-based model, which is a non-parameter model for financial asset pricing; the other is the rule-based model, which focuses on the profitable trading rule discovery.

When constructing the pricing-based model, the fluctuation of the financial asset price is forecast by approximating the functional mapping between the financial asset price and its influencing factors. The market price in the next trading period is forecast using historical financial time series data and relative factors such as technique analysis indicators or economic indicators.

Fuzzy rule is commonly used for stock price prediction model and can be implemented in a real-time trading system. Chang and Liu (2008) proposed a Takagi-Sugeno-Kang (TSK)-type fuzzy rule-based system for stock price prediction. Furthermore, this model is improved by combining the wavelet transform (Chang & Fan, 2008). Zarandi, Rezaee, Turksen, and Neshat (2009) proposed a type-2 fuzzy rule-based expert system model for stock price analysis. Although these models can obtain more accurate prediction results than the traditional regression model, they only consider the investment trading, and they can hardly be applied for arbitrage trading.

Some arbitrage trading systems have been proposed based on the target price forecasting. An index arbitrage model for the Irish market index (ISEQ) and the FTSE 100 index is demonstrated using recurrent neural network models combined with the Kalman filter (Edelman, 2008). An option arbitrage trading system for American-style call options on the British Pound versus the US dollar currency futures is proposed based on a novel pseudo self-evolving cerebellar model arithmetic computer (PSECMAC) option-pricing model (Teddy et al., 2008). However, these research only demonstrate the detection for daily arbitrage trading and do not show capability when applied to practical intra-day arbitrage trading

When constructing the rule-based model, the trading rule is commonly generated by identifying the charting patterns. A new template grid, which is a matching technique based on pattern recognition, is proposed to detect bull flag technical trading rules (Wang & Chan, 2007) and buy signals (Wang & Chan, 2009). Li and Kuo (2008) combined the K-chart technical analysis, wavelet transform, and self-organizing map network to construct a forecasting model and to generate buying and selling signals.

More recently, the hybrid models, which combine the price forecasting model and rule discovery mechanism, have been proposed in many literature. Tan, Quek, and Yow (2008) proposed a novel rough set-based pseudo outer-product (RSPOP) fuzzy neural network intelligent stock trading system. They combined the price predictive model and technique indicator predictive model to obtain the optimal trading rules. Ghandar, Michalewicz, Schmidt, To, and Zurbrugg (2009) used an evolutionary process in trading rules drawn from a fuzzy logic rule base. However, no literature has focused on inter-market arbitrage trading.

In order to develop a profitable and easy to implement system for arbitrage trading, we combine expert knowledge and statistical analysis to determine the properly trading timing conditions. Moreover, we then use these conditions to filter rapidly the transaction data stream and reduce the computational loading for high-frequency data processing. Furthermore, the knowledge discovery process is applied to generate the arbitrage trading rule using the filtered data.

XCS is a knowledge discovery process that has already been applied in various related studies on financial investment, and it has shown the capacity to process financial time series data. These studies also indicate that the XCS model is more profitable compared with the random or buy and hold model (Beltrametti et al., 1997; Liao & Chen, 2001; Schulenburg & Ross, 2002). Therefore, this study will utilize the classifier system's dynamic learning function to build a self-learning, self-adaptive inter-market arbitrage model that can be applied within a dynamic market. Using the SGX MSCI Taiwan Index Futures and TAIFEX TX as the study object, we will assess the effect of actual costs of trading, price spreads, different expiry dates, and trading timing, and then design an inter-market arbitrage investment decision support system that can be put to practical use.

The rest of the paper is organized as follows: Part 2 illustrates the inter-market arbitrage strategy and method in this study; Part 3 describes the proposed XCS model; Part 4 details the experiment design and the results; and lastly, conclusions drawn from the study are discussed in Part 5.

2. Arbitrage Strategy Analysis

2.1. Inter-market arbitrage

When conducting futures index arbitrage, the common method is to calculate the theoretical fair price based on the term to maturity and the underlying commodity first and then compare it to the actual market price to calculate the extent of mispricing. To assess the price spread inter-market, the extent of mispricing must be normalized as follows:

$$M_t^i = \frac{F_t^i - FP_t^i}{FP_t^i} \tag{1}$$

where M_t^i is the normalized mispricing at time *t*, F_t^i is the actual future price, FP_t^i is the theoretical fair price of the index futures, and *i* is the individual futures contractor.

The mispricing differential between the two related products is used as the inter-market price spread. Using the indices nominated for study the SGX MSCI Taiwan Index Futures and the TAIFEX Taiwan stock index futures would be as follows:

$$SM_t = M_t^T - M_t^S \tag{2}$$

where SM_t is the spread mispricing differential, and M_t^T and M_t^S represent the normalized mispricing in the TAIFEX Taiwan Stock Index Futures and the SGX MSCI Taiwan Index Futures, respectively.

When SM_t reaches a certain extent and becomes greater than the cost of the arbitrage transaction (TC), that is $|SM_t| > TC$, then proceeding with the arbitrage trading is worth considering. Using Eq. (2) as an example, due to the special characteristic of the index futures price converging towards the spot price as the expiry date for the futures contract is approaching, where spread was found to be underpriced ($SM_t < 0$), an arbitrageur would take a long position in the TAIFEX Taiwan stock index futures contracts and simultaneously set up an opposing short position in the MSCI Taiwan Index Futures contracts.

In general, the cost of carry model is most commonly applied to calculate the theoretical fair price of the index futures using the Eq. below:

$$FP_t^i = I_t^i \cdot e^{(r_i - d_i)(T_i - t)} \tag{3}$$

where I_t^i equates to the price of the underlying stock index at time *t*, r_i refer to the risk-free interest rate, d_i is the dividend yield rate, and T_i is the futures contract expiration date.

To process high frequency inter-minute data in inter-market arbitrage trading, we simplified Eq. (3) by replacing FP_t^i with I_t^i and then substituting this in Eq. (1) and (2) to calculate M_t^i and SM_t. Although, the expiry date effect in Eq. (3) is neglected, it is still considered as an important factor when designing XCS in this study. The modified Equations in the study are expressed as follows:

$$\tilde{M}_t^i = \frac{F_t^i - I_t^i}{I_t^i} \tag{4}$$

$$\tilde{S}\tilde{M}_t = \tilde{M}_t^T - \tilde{M}_t^S \tag{5}$$

2.2. Determining arbitrage trading conditions

To increase the efficiency of processing high frequency data from inter-minute trading, we search for the conditions that can be applied to data filtering, based on the characteristics of historical daily trading data, and the magnitude of the probability of successful arbitrage.

First, we use Eq. (4) to calculate respectively two futures products' inter-market trading price spread between each futures product's price and spot price. We then substitute this into Eq. (5) to calculate whether the absolute value obtained is greater than the cost of transactions and then proceed with arbitrage trading.

Therefore, we can conduct classification based on the different combinations of scenarios with price spreads correlation to calculate the probability of success of arbitrage, using it to discover the most frequent condition for successful arbitrage under various correlated price spread combinations. The conditions are listed in Table 1 and illustrated in Fig. 1.

Table 1. Condition classification for statistical arbitrage

No. condition	$\tilde{M}_t^T$	$\tilde{M}_t^S$	$\tilde{S}\tilde{M}_t$	**Trading position**	
				TX	**MCSI**
1	> 0	> 0	> 0	Short	Long
2	> 0	> 0	< 0	Short	Long
3	> 0	< 0	> 0	Short	Long
4	< 0	< 0	> 0	Long	Short
5	< 0	< 0	< 0	Long	Short
6	< 0	> 0	< 0	Long	Short

The probability of successful arbitrage for each condition was then calculated using the support value widely used for mining association rules (Han, 2006).

$$\begin{aligned} & support = probability\ of\ successful\ arbitrage \\ & = \frac{\text{total no. of times of profit - making}}{\text{total no. of occurrence of the conditions}} \end{aligned} \tag{6}$$

Finally, the conditions with high support value would be used for filtering arbitrage opportunities. Only the data that matches the condition would be used for the XCS knowledge discovery process.

2.3. Trading interval and portfolio management

When constructing an inter-market arbitrage trading model, we must first set an entry interval in arbitrage trading based on the respective expiry dates for the product in each market. A the same time, this interval must be similar to the longest holding period for the futures, so as to avoid the effects caused by the different portfolio management in practice due to any difference in expiry dates, and to avoid the occurrence of closing out or rolling over of position at expiry. We set the arbitrage trading interval using the day after the MSCI Taiwan Index Futures contract settlement date and the following month's last trading day of the TX contract as one period. The last exit trading day is the TX contract settlement date. The trading interval is illustrated in Fig. 2.

Furthermore, to calculate the final profit/loss from arbitrage trading, establish the relationship between the corresponding prices inter-market, and avoid the assumption of too much trading risk via the establishment of stop-loss, profit-cap mechanisms, when the profit from trading is greater than the profit-cap threshold, profit-realization will be conducted. Conversely, if the loss exceeds the stop-loss threshold, immediate action will also be taken to exit and minimize loss. In addition, if during the trading period, neither the stop-loss nor the profit-cap trading is triggered, then on the last trading day for the TX contract, exit trading will be executed at the spot index price of the last order (at 13:30) and profit/loss will be calculated.

This study assumes relevant trading parameters based on practical experiences, which are set out as follows:

- Price spread ratio (hedge ratio): TX contracts: MSCI Taiwan Index Futures contracts = 3:4

- Transactions costs: calculating retail transaction fees and tax, the total cost of trading is approximately TWD5,800.

- Stop-loss & Profit cap: this study utilizes the knowledge acquisition training period, uses the loss-making (profit-making) investment trading data as a statistical sample to calculate the distribution of dollar value lost (profit), and sets the stop-loss (profit-cap) value to cut loss (profit) at 30% (70%) of the maximum loss (profit). This is illustrated in Fig. 3 below.

3. The XCS-based Model for Arbitrage

3.1. Extended classifier system

The classifier system is an adaptive rule-base system consisting of enhanced learning mechanisms and the genetic algorithm, which is capable of developing various combinations of rules within the system to acquire optimal rules. Therefore, the classifier system can categorize external states, accurately yield predictions, and can also adapt to changes in external states, thereby generating different predictions under different states to reflect the appropriate solution applicable to the dynamic environment.

The original concept of the classifier system came from Holland (1976), under the term Cognitive System (CS). Following, Holland and Reitman (1977) jointly proposed the Learning Classifier Systems (LCS). Since then, subsequent research conducted by many scholars gradually strengthened the overall operational efficiency and stability of the system.

An improved version of the learning mechanism was proposed by Wilson (1995, 1998). He adjusted the fitness of LCS, changing the original use of expected return as a basis for calculating the accuracy of the expected return. He also improved the algorithm for learning. The improved model was named Extended Classifier Systems (XCS).

In XCS, the so-called classifier is composed of many “IF condition/ THEN action” rules to represent the corresponding external state. This is represented by the following formula:

<classifier>：= <condition>/<action> (7)

For the sake of easy application, binary coding is typically used for the condition and the action to represent various parameters of the external state. It is also used as a code for the following set of instructions:

<condition>：= {0,1,1#,0,1,….}L (8)

<action>：= {0,1,….,n-1} (9)

Within these codes, L represents the length of the rules, # represents the unimportant characteristics which mean that 0 and 1 can both be matching states, and n represents the classified resulting numbers.

The main structure and application process are represented in Figure 4. The algorithm of the XCS model is shown in Figure 5.

As can be seen, XCS receives information on the external state through detectors, coding it into chains of rules that can be processed by the system. These chains of rules are called classifiers. These classifiers are then compared to the classifiers identified in the external state's information system and population set [P], and those that match the current imputed state are selected to create a match set [M]. If no matching classifiers are found in the population set, then the cover mechanism is triggered to set up one that contains the set of information as that point in time, and action will be randomly generated thereafter. From the action of each classifier in the match set, the weighted average of each action is then calculated based on the fitness of the classifiers to construct a prediction array [PA] for returns. Finally, the appropriate action is determined through the random exploration or exploitation method. This action is then used to set up an action set [A]. After determining the appropriate action, the system delivers the action to the effector to be sent for execution under the given conditions. Depending on the level of correctness resulting from the execution, the system will then provide internal reinforcement to the classifiers, and the relevant weighting in terms of the strength of each classifier within the action set is thus updated. Afterwards, the evolutionary genetic algorithms mechanism is applied within the action set, which will then eliminate the relatively weak rules. Therefore, after a period of learning, the system can generate the most appropriate action classifier that can adapt to the various states created by various changes within a dynamic environment.

3.2. The proposed xcs-based arbitrage model

This study uses XCS to establish an inter-market arbitrage model. Fig. 6 represents the structure of the arbitrage trading system in this study, which consists of three main components: data processing, XCS, and portfolio management.

First, in order to deal with the huge volume of intra-day 1-minute transaction data, in the data pre-processing stage, the arbitrage conditions are determined using the association rule mining approach described in Section 2.2. The price spread correlations, i.e. the positive or negative sign of $\tilde{M}_t^T, \tilde{M}_t^S$ and $\tilde{S}\tilde{M}_t$, are applied to calculate the most frequent item and then to filter out high-risk arbitrage opportunities so as to increase the efficiency of the model's high frequency data operation.

Second, with the remaining data, the XCS learning algorithm is applied to conduct a purification of trading knowledge to find the descriptive factor for the most suitable state for arbitrage trading. XCS is a type of self-adjusting learning algorithm. Based on the dynamic factors of the state, it can search for conditions with the highest fitness. Therefore, we have chosen some of the most commonly observed market trading data to be used as descriptive factors for setting up the conditions for arbitrage. These factors are imputed into XCS and tied in with calculations of the return in value for arbitrage to facilitate learning of hidden knowledge and to seek applicable knowledge and rules for arbitrage trading.

Finally, in the portfolio management component, we take into account some of the demands of trading in practice, such as stop-loss, profit cap, and closing out of positions at expiry. Daily settlement for profits in arbitrage trading is conducted and applied to manage the arbitrage portfolio, eventually working out the investment decision of buying, holding, or selling.

3.3. Knowledge encoding and discovery by XCS

In XCS, the so-called classifier is made up of many ‘IF condition/ THEN action’ rules to represent the corresponding external state. Usually, for the sake of easy application, binary coding is used for the condition and action to represent various parameters of the external state.

Based on Eq.s (4) and (5) described in Section 2.1, we use the price spread and the price spread ratio as the conditions descriptive factors of the classifier system. Further, the time of trading is also an important factor that will influence the arbitrage profitability. The expiry date are considerd as the conditions descriptive factors of the classifier system. The performance of arbitrage activities is different within the same trading day and a certain time in the day tends to be favored (Taylor, 2007). To take into account inter-day activities, we also add in intra-day transaction times as a state descriptive factor.

These conditions and action of the classifier must go through a process of discretization before binary coding can be conducted. Therefore, we use a linear function to conduct discretization. Table 2 shows the composition of the classifier in this study.

Table 2. Composition of the classifier

Conditions				**Action**
Bit 1-4	**Bit 5-8**	**Bit 9-12**	**Bit 13-14**	**Bit 15**
price spread between TX and MSCI ($\tilde{M}_t^T$)	price spread ratio of TX ($\tilde{S}\tilde{M}_t$)	the term to expiry date for TX ($T_i - t$)	Intra-day trading timing	Profitable (True/ False)
Examples: 0<X<= 3 /16 →0 3/16<X<= 6 /16 →1 …… 15/16<X →15	0 <Y<= 2 /16 →0 2/16<Y<=6 /16 →1 …… 30/16<Y →15	Z = 1 →0 Z = 2 →0 …… Z = 16 →15	9:00<=T<10:00→0 10:00<=T<11:00→1 11:00<=T<12:00→2 12:00<=T<13:30→3	True (profit) → 0 False (loss) → 1

Moreover, in the process of knowledge discovery, there are some parameters for the XCS operation that need to be defined. In this study, the pay-off function used in XCS follows the hierarchical structured proposed by Wilson (1998). The three levels set are maximum profit (loss), stop-loss (profit cap), and minimum profit (loss). When XCS is in operation, the evolution of classifier rules are also generated by the genetic algorithm used for setting the relevant parameters according to the best value set as proposed by Wilson (1998).

3.4. Evaluation method

This study follows the prediction of the XCS arbitrage model, using intra-day 1-minute trading data to conduct testing. Based on trading time and trading data, predictions are made on whether or not inter-market arbitrage trading can be conducted. We measure the accuracy of prediction and profitability respectively and use them as indicators to evaluate the model. They are defined as follows:

1. Accuracy:

$$Accuracy = \frac{\text{no. of times of actual profit generation}}{\text{no. of transactions}} \tag{10}$$

2. Profitability:

$$\begin{aligned} &Trading\ profit/loss \\ &= (\text{sale price}) - (\text{purchase price}) - (\text{transactions costs}) \end{aligned} \tag{11}$$

$$\begin{aligned} &Profitability = Average\ profit/loss\ per\ trade \\ &= \frac{\sum Trade\ profit/loss}{\sum number\ of\ transactions} \end{aligned} \tag{12}$$

4. Empirical Result

4.1. Data and experiment design

This study obtains the empirical trading data from the TAIFEX Taiwan Index Futures (spot month), the Taiwan Weighted Index, the MSCI Taiwan Index Futures (spot month), the MSCI Taiwan Index, and the corresponding underlying index's intra-day 1-minute data. The foreign exchange rate of TWD/USD is the daily close price. All the transaction data are provided by the APEX International Financial Engineering. Empirical analysis is undertaken on data obtained within the intervals and selected from January 1, 2001 to December 30, 2006. Due to the fact that this study uses price spreads as the benchmark, when executing strategies, the timing used is mainly the intersection between the spot market's and the futures market's opening times. The intra-day trading period for the sample data collection is between 9:00 in the morning and 13:30 in the afternoon.

The main aim of designing the experiments is to look closely at the XCS model's applicability in arbitrage and compare it against a random trading strategy. That is, when the XCS arbitrage model determines at a certain point in time whether or not to give prediction to an arbitrage opportunity, the random model would also generate a random trading signal (which corresponds to an action generated from the XCS model). It would also execute arbitrage trading according to the random trading signal generated, holding the positions till expiry without any allowance for any stop-loss or profit-cap mechanism.

To further test the effects of different characteristics of the data collected from different calendar years in trading and different time intervals in the XCS model, three types of training and testing datasets are designed, each undergoing empirical simulation using the XCS model and the random model to conduct a total of six types of testing. The experiment design and the relevant variants collation are shown is Table 3.

Table 3. Experiment design

Testing model	No. dataset	Time-frame	Training period	Year of testing
(1) XCS (2) Random	DS 1	4 years	2001~2003	2004
	DS 2	5 years	2001~2004	2005
	DS 3	6 years	2001~2005	2006

4.2. Arbitrage conditions for data stream filter

To reduce the data size, processing and filtering of one minute high frequency data are one of the key processes for arbitrage in this study. Meanwhile, the correlation between price spread and successful arbitrage provides a benchmark for the situational description classification in our model's assessment of arbitrage success. Following the method described in Section 2.2 and substituting the daily trading data collected between January 1, 1998 and December 30, 2006 into the model, the resulting conditions and their support values are listed in Table 4 below.

From the table, it can be seen that Conditions 3 and 6 yield better arbitrage opportunities. Therefore, this study utilizes these two rules as data filtering conditions. Only intra-day 1-minute data that fit these two conditions will be entered into XCS for knowledge discovery.

Table 4. Arbitrage conditions determination

No. Condition	$\tilde{M}_t^T$	$\tilde{M}_t^S$	$\tilde{S}\tilde{M}_t$	Support
1	> 0	> 0	> 0	150/295 (0.51)
2	> 0	> 0	< 0	60/270 (0.22)
3	> 0	< 0	> 0	**95/157 (0.61)***
4	< 0	< 0	> 0	60/237 (0.25)
5	< 0	< 0	< 0	81/203 (0.51)
6	< 0	> 0	< 0	**112/197 (0.57)***

4.3. Knowledge rules analysis

In this study, we performed a preliminary experiment to illustrate the knowledge discovery ability of the XCS model. Simulateously, we simulated the arbitrage trading with intra-day 1-minute data to illustrate the trading system.

In the preliminary expriment, we set the arbitrage condition for filtering the data stream as Condition 3 and the dataset for expriment as DS 3. The XCS-based arbitrage model was trained and tested according to the intra-day 1-minitue trading data for six years. During model training, only 26,471 pieces of 1-minute trading data distributed in 395 trading days from year 2001 to year 2005 could fit the condition and pass the data stream filter. After XCS training, 227 trading rules were generated from 26,471 pieces of 1-minute trading data, which were then used for testing.

During testing, only 1,108 pieces of 1-minute trading data in the year 2006 could fit the condition and pass the data stream filter. Among these, only 359 pieces of data were matched with 57 rules, which were generated by XCS. We then excuted the arbitrage transaction. We captured the trading situation of one trading day in the testing period to illustrate the capability of the XCS model.

Figure 7 is the simulated trading situation on 2 Janurary 2006. Small pieces of the streaming data during the whole trading day were filtered out as the arbitrage interval. Among the 271 pieces of intra-day 1-minute trading data from 9:00 AM to 13:30 PM, only 46 pieces of data could fit Condition 3, and they were viewed as the arbitrage opportunities. Seventeen pieces of 1-minute trading data among the arbitrage opportunities were matched with the XCS rules, and the arbitrage transaction was then executed. In Figure 7, we can observe that the arbitrage opportunities were distrbuted during the period from market opening at 9:00 AM to 10:27 AM. The arbitrage transactions were concentrated between 9:17 AM and 9:52 AM, lasting about five minutes. The arbitrage opportunities did not occur every time. Only a few arbitrage opportunities matched the XCS rules and were suitable to excute abitrage transaction.

In order to understand the meaning of the knowledge rule generated by XCS, the top five rules selected by the correctness rate and occurrence times in training and testing are listed in Table 5.

Table 5. Knowledge rules generated by XCS-based arbitrage model

Rules		**Training**		**Testing**	
No.	**Condition**	**Cor. rate**	**Occ. times**	**Cor. rate**	**Occ. times**
Top 5 rule of correctness rate in training					
87	10000110101	100 %	41	N. A.	0
100	00000010111	100 %	36	78 %	9
130	00000000010	100 %	34	0 %	13
98	00000010100	100 %	25	N. A.	0
178	10000010001	100 %	21	100 %	3
Top 5 rule of correctness rate in testing					
200	00000010010	100 %	1	100 %	28
216	10000000001	100 %	13	100 %	25
196	00000010000	75 %	4	100 %	25
201	10000010011	65 %	17	100 %	19
153	10000111110	100 %	3	100 %	17
Top 5 rule occurred in training					
46	10000000011	61%	62	100 %	1
48	00000000110	77 %	47	100 %	6
87	10000110101	100 %	41	N. A.	0
27	00000000011	95 %	40	30 %	10
28	10000101111	97 %	38	N. A.	0
Top 5 rule occurred in testing					
200	00000010010	100 %	1	100 %	28
216	10000000001	100 %	13	100 %	25
196	00000010000	75 %	4	100 %	25
201	10000010011	65 %	17	100 %	19
153	10000111110	100 %	3	100 %	17

In Table 5, we can observe that the knowledge rules with high correctness rate in training seldom occurred during testing; similary, the knowledge rules with high correctness rate in testing seldom occurred during training. The highest correctness rate of the knowledge rules both in training and testing were 100%. However, the correctness rate of the knowledge rules that occurred most frequently during training was less than 100%. Moreover, the knowledge rules with the highest correcness rate were consistent with the rules that occurred most frequently during testing.

Based on the above analysis, we conclude that for a certain knowledge rule generated by XCS, the correctness rate in training is inconsistent during testing. We also find that the rule with the highest correctness rate in training does not work during testing, while the rule with the highest correctness rate in testing does not work as well as that during training. However, the rules that occurred most frequently and have the highest correctness rate are consistent during testing. Applying these rules in trading can help gain profit.

4.4 Model comparison

From the two conditions derived in the previous section, Conditions 3 and 6 are used as the filtering conditions. The XCS model then undergoes a processes of learning and verification, testing the accuracy and profitability of the model. As the XCS model uses the genetic algorithm (GA) for selection of the fittest situational factor and has a random mutation characteristic, repeated experiments on the same dataset will yield results that are inconsistent with the expectations. Repeated experiments yielding inconsistent results can also occur in the random trading model. For this reason, the experiment on the same dataset should be repeated for 10 trials, and the average and standard deviation of these 10 sets of experiment results should be calculated. The results of the experiment under the two conditions are collated in Table 6 and Table 7.

Table 6. Model comparison of accuracy

Model	Dataset	Accuracy (%)			
		Min.	Max.	Ave.	Std.
Condition 3					
XCS	DS 1	59.28	77.45	69.58	6.79
	DS 2	48.23	53.73	52.13	1.65
	DS 3	74.86	83.25	79.29	3.16
Rand.	DS 1	47.15	49.68	48.53	0.86
	DS 2	47.22	49.21	48.55	0.6
	DS 3	45.48	50.63	48.67	1.58
Condition 6					
XCS	DS 1	65.78	68.41	66.45	1.05
	DS 2	27.65	34.61	31.87	2.12
	DS 3	66.05	75.87	70.55	3.13
Rand.	DS 1	47.02	49.19	48.37	0.73
	DS 2	47.83	49.63	48.68	0.59
	DS 3	48.41	49.42	48.80	0.39

Table 7. Model comparison of profitability

Model	Dataset	Profitability (per trading)			
		Min.	Max.	Ave.	Std.
Condition 3					
XCS	DS 1	2674	15805	8538	4462
	DS 2	89	1782	889	496
	DS 3	10187	13752	12025	1217
Rand.	DS 1	-1241	2281	-38	1048
	DS 2	-288	422	147	272
	DS 3	-3498	11904	1740	4912
Condition 6					
XCS	DS 1	8538	10259	9462	639
	DS 2	-7794	-4814	-5906	577
	DS 3	4785	16059	10140	3369
Rand.	DS 1	-1525	1756	-137	1013
	DS 2	-842	424	-235	360
	DS 3	-581	845	-17	430

From Table 6, it can be observed that in terms of accuracy, through the inter-market arbitrage model’s constructed random model, there is close accuracy in all the simulated trading scenarios. However, the accuracy of the XCS model compared with all the simulated trading scenarios is distinct. For the XCS model, apart from the average accuracy of Condition 6 not exceeding 50% during the testing period in 2006 (DS 2) and having the lowest value of 31.87% amongt all the experiments, its other situational accuracy levels all exceed 50% and are superior to the random model (which averaged around 48% in accuracy across all situations). Amongt the experiments, Condition 3 generated the maximum average accuracy of 79.29% during the testing period in 2006 (DS 3).

From Table 7, in terms of profitability, regardless of condition and dataset selection, all tests of the XCS model – with the exception of Condition 6 in the 2006 (DS 2) testing period which results in a negative return and underperformed the random model – has shown far greater profitability than the random model.

5. Conclusion

In inter-market arbitrage, if two futures products have different expiry dates, then a closing out of position will be forced upon the underlying commodity that first reaches expiry, exposing other components that are yet to expire due to risks. This will in turn lead to the failure of the arbitrage strategy. These risks are difficult to quantify using traditional financial engineering. Therefore, this study addressed this issue by proposing a dynamic learning, adjustable XCS inter-market arbitrage trading model to reduce risks associated with inter-market arbitrage. In addition, this study also proposed a solution for handling large volumes of intra-day 1-minute trading data. By using association rules to filter high frequency data, inter-market arbitrage opportunities can be immediately identified through searching the intra-day one minute trading data.

This study used nearly six years' of intra-day 1-minute trading data to conduct this empirical research, measuring the XCS model's accuracy and profitability and comparing it with the testing results generated by random trading strategies. Results from this research show that – compared to the random model – by using factors such as price spread ratio, expiry date, and intra-day trading time to build the XCS inter-market arbitrage model, it yields sufficient accuracy and profitability and can effectively lower the risks associated with inter-market arbitrage.

References

Beltrametti, L., Fiorentini, R., Marengo, L., & Tamborini, R. (1997). A learning-to-forecast experiment on the foreign exchange market with a classifier system. *Journal of Economic Dynamics & Control, 21*(8-9), 1543-1575.

Butterworth, D., & Holmes, P. (2002). Inter-market spread trading: Evidence from UK index futures markets. *Applied Financial Economics, 12*(11), 783.

Chang, P., & Fan, C. (2008). A hybrid system integrating a wavelet and TSK fuzzy rules for stock price forecasting. *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, 38*(6), 802-815.

Chang, P., & Liu, C. (2008). A TSK type fuzzy rule based system for stock price prediction. *Expert Systems with Applications, 34*(1), 135-144.

Cornell, B., & French, K. R. (1983). The pricing of stock index futures. *Journal of Futures Markets, 3*(1), 1-14.

Dunis, C. L., Laws, J., & Evans, B. (2006). Trading futures spreads: An application of correlation and threshold filters. *Applied Financial Economics, 16*(12), 903.

Dunis, C. L., Laws, J., & Evans, B. (2008). Trading futures spread portfolios: Applications of higher order and recurrent networks. *The European Journal of Finance, 14*(6), 503.

Edelman, D. (2008). Using kalman-filtered radial basis function networks for index arbitrage in the financial markets. *Natural Computing in Computational Finance; STUDIES IN COMPUTATIONAL INTELLIGENCE; 1st European Workshop on Evolutionary Computation in Finance and Economics,* Valencia, SPAIN. , *100* 187-195.

Figlewski, S. (1984). Explaining the early discounts on stock index futures: The case for disequilibrium. *Financial Analysts Journal, 40*(4), 43-47.

Ghandar, A., Michalewicz, Z., Schmidt, M., To, T., & Zurbrugg, R. (2009). Computational intelligence for evolving trading rules. *Ieee Transactions on Evolutionary Computation, 13*(1), 71-86.

Han, J. & Kamber, M. (2006) *Data Mining: Concepts and Techniques (2nd ed.),* Morgan Kaufmann,

Holland, J. H. (1976). Adaptation. In Rosen, R. and Snell, F. M., editors, *Progress in Theoretical Biology,* 4. Plenum, New York.

Holland, J. H., & Reitman, J. S. (1977). Cognitive systems based on adaptive algorithms. *SIGART Bull.*, (63), 49-49.

Klemkosky, R. C. & Lee, J. H. (1991). The intraday ex post and ex ante profitability of index arbitrage. *Journal of Futures Markets, 11*(3), 291-311.

Li, S., & Kuo, S. (2008). Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks. *Expert Systems with Applications, 34*(2), 935-951.

Liao, P. Y., & Chen, J. S. (2001). Dynamic trading strategy learning model using learning classifier systems. *Proceedings of the 2001 Congress on Evolutionary Computation, Vols 1 and 2; IEEE Congress on Evolutionary Computation; Congress on Evolutionary Computation (CEC 2001),* SEOUL, SOUTH KOREA. 783-789.

Luo, W. C. (2002), Spread Arbitrage between Stock Index Futures in Taiwan: A Cointegration Approach, Working paper, University of Southampton.

Modest, D. M., & Sundaresan, M. (1983). The relationship between spot and futures prices in stock index futures markets - some preliminary evidence. *Journal of Futures Markets, 3*(1), 15-41.

Moles, P. & Terry, N. (1997), *The Handbook of International Financial Terms*, Oxford University Press.

Montana, G., Triantafyllopoulos, K., & Tsagaris, T. (2009). Flexible least squares for temporal data mining and statistical arbitrage. *Expert Systems with Applications, 36*(2), 2819-2830.

Schulenburg, S., & Ross, P. (2002). Explorations in LCS models of stock trading. *Advances in Learning Classifier Systems; LECTURE NOTES IN ARTIFICIAL INTELLIGENCE; 4th International Workshop on Learning Classifier Systems,* SAN FRANCISCO, CALIFORNIA. *, 2321* 151-180.

Tan, A., Quek, C., & Yow, K. C. (2008). Maximizing winning trades using a novel RSPOP fuzzy neural network intelligent stock trading system. *Applied Intelligence, 29*(2), 116-128.

Taylor, N. (2007). A new econometric model of index arbitrage. *European Financial Management, 13*(1), 159-183.

Teddy, S. D., Lai, E. M. -K., & Quek, C. (2008). A cerebellar associative memory approach to option pricing and arbitrage trading. *Neurocomputing; 13th International Conference on Neural Informational Processing,* Hong Kong, PEOPLES R CHINA. *, 71*(16-18) 3303-3315.

Wang, J., & Hsu, H. (2006). Degree of market imperfection and the pricing of stock index futures. *Applied Financial Economics*, *16*(3), 245.

Wang, J., & Chan, S. (2007). Stock market trading rule discovery using pattern recognition and technical analysis. *Expert Systems with Applications*, *33*(2), 304-315.

Wang, J., & Chan, S. (2009). Trading rule discovery in the US stock market: An empirical study. *Expert Systems with Applications*, *36*(3), 5450-5455.

Wilson, S. W. (1995). Classifier fitness based on accuracy. *Evol.Comput.*, *3*(2), 149-175.

Wilson, S. W. (1998). Generalization in the XCS classifier system. In J. R. Koza et al. (Eds.), *Genetic Programming 1998: Proceedings of the Third Annual Conference*, pp. 665-674. Morgan Kaufmann.

Zarandi, M. H. F., Rezaee, B., Turksen, I. B., & Neshat, E. (2009). A type-2 fuzzy rule-based expert system model for stock price analysis. *Expert Systems with Applications*, *36*(1), 139-154.

計畫成果自評

融合金融與資訊的創新領域金融工程物理學在此展開學新的一頁，改寫過去金融資訊的歷史，正式的定義以過去的歷史資料尋找出交易資料間的力道表現，一如物理學般的表現，當市場受到動力推動時，漲者恆漲，跌者恆跌。透過發表的成果可以得知，利用資訊技術搜尋這樣的力道是可行且被學者認可。根據此理論所設計的模型也能應用於實際的金融投資領域上，透過資訊技術建構的模型搜尋市場的力道與異常行為，並且預測市場接下來可能發生的行為，最終的目標在於使投資人能夠在投資市場中減少虧損，增加獲利。投資人最害怕的就是資訊的缺乏，然而金融環境永遠不會有資訊充足的狀況，因此有些人能夠比其他人獲得更多資訊，進而先行一步制訂策略，造成其他投資人的損失。因此透過歷史資料學習隱藏在市場資料中的關鍵機會，例如力道的轉移，或者是盤勢的改變行為等各種市場行為，根據金融工程物理學的模型提供預測資訊，能盡量避免投資決策制訂錯誤。金融工程物理學不但能將金融市場行為進行解釋，同樣的也能應用在投資心理行為的解釋，因為此領域強調的在於因果的發現，其後所能應用的範圍也可拓展至資產組合、避險基金管理等應用面。

本計畫針對動態的環境進行資產配置與投資操作，實能將金融工程物理學應用於此，根據整體研究成果而論，動態模型及計算智慧能夠適應環境劇烈的變動，並制訂符合當時情況的策略，使得投資獲利能夠大幅提昇，在模型的準確率也能提高許多，不只改善傳統理論的不足之處，更能創新模型。本研究發表於多個期刊及研討會上(見相關著作)，更證明此一領域具備高度實務及學術價值。

在本計畫執行期間，所有成果皆展示於金融投資決策教學研究中心，此教學中心提供國內外學生及學者進行參訪及使用。本中心慕名前來參觀學者非常繁多，包含 Oklahoma state university、California State University、上海師範大學、北京大學等校之學者。透過這些多次國際交流經驗也瞭解到本研究領域在世界上是保持領先，使得各國學者皆爭相參觀，期望能學習到最新的知識。

本人也曾多次在國內多所大學及中國大陸進行課程教學，並將此知識進行深度的探討與傳授給學子，課程結束後也獲得學子們的熱烈歡迎，並進而實作了相當多的應用，使本領域的知識逐漸擴散。

相關著作發表

1.第一年

1. Yu-Chia Hsu, An-Pin Chen (2008), "Clustering Time Series Data by SOM for the Optimal Hedge Ratio Estimation," Proceedings of International Conference on Convergence Information Technology (ICCIT '08), Nov. 11-13, 2008. Busan, Korea. (EI)

2. Tsai Wen-Chih, An-Pin Chen (2008), "Global Asset Allocation using XCS Experts in Country-Specific ETFs," Proceedings of International Conference on Convergence Information Technology (ICCIT '08), Nov. 11-13, 2008. Busan, Korea. (EI)

3. Yu-Chia Hsu, An-Pin Chen (2009), “SOM-Based Hedge Ratio Estimation with Hierarchical Cluster Resampling,” Proceedings of the 12th IEEE International Conference on Computational Science and Engineering (IEEE CSE'09), August 29-31, 2009, Vancouver, BC, Canada. (EI)

2.第二年

1. Yu-Chia Hsu, An-Pin Chen, Jia-Haur Chang (2010), “An inter-market arbitrage trading system based on extended classifier systems.” Expert Systems with Applications, doi:10.1016/j.eswa.2010.09.039. (forthcoming), (SCI, IF=2.908)

2. An-Pin Chen, Yu-Chia Hsu (2010), “Dynamic Physical Behavior Analysis for Financial Trading Decision Support,” IEEE Computational Intelligence Magazine, Vol. 5, issue 4, 2010. (forthcoming) (SCI, IF=2.54)

3. Wen-Chih Tsai, An-Pin Chen(2010), “Strategy of global asset allocation using extended classifier system. Expert Systems with Applications”, pp 6611-6617, Volume 37 , Issue 9, 9 March 2010 (SCI, IF=2.908)

4. Wen-Chih Tsai, An-Pin Chen(2010), “Using the XCS Classifier System for portfolio allocation of MSCI index component stocks”. Expert Systems with Applications, pp 6611-6617, Volume 37 , Issue 9, 9 March 2010 (SCI, IF=2.908)