



# Technical Trading Behaviour: Evidence from Chinese Rebar Futures Market

Guanqing Liu<sup>1</sup>

Accepted: 1 September 2018 / Published online: 10 September 2018  
© The Author(s) 2018

## Abstract

Technical Traders adopt mathematical methods to formulate various technical trading rules on their trading strategies. This paper utilises two unique datasets—individual and market tick-by-tick data—to disclose the categories and characteristics of technical traders' strategies in Chinese rebar futures market. Firstly, we use a simple multiple regression model to filter technical traders in individual dataset. By using market dataset to generate dummy signals according to six popular kinds of technical rules, we created dummy trading directions as benchmark for real trading actions. Based on the similarity between dummy signals with different technical rules and traders' real actions, we employ k-means algorithm to classify technical traders. Through these empirical works, technical traders in my dataset are classified into 11 groups. Finally, on the basis of 11 clusters' coordinates, the features of technical strategies in each group are summarised.

**Keywords** Rebar futures market · Technical trading · K-Means clustering

## 1 Introduction

How could we possibly disclose human behaviour in any part of society is an impossible fully answered question in the world, because we are not “unicellular” animals and brain control us to make uncountable behaviours. For this work, we could do something relevantly in financial market which is a special and important part of society. We suppose to classify market participants and observe their trading behaviours to find any patterns. A special kind of market participants are created and controlled

---

Dr. Guanqing Liu graduated from University of Leicester with Ph.D in Economics in 2016.

✉ Guanqing Liu  
guanqing\_liu@yahoo.com

<sup>1</sup> Department of Economics, University of Leicester, Leicester, UK

by human themselves—technical traders. This paper focus them and investigate their behaviour possibly.

Generally, trader's types have many kinds of identification, and these diversities could reflect on the trading behaviour. Current theoretical researches on market microstructures usually divide traders into informed and uninformed traders (Bagehot 1971). With development of the theory of financial market microstructure, traders also are classified as patient and impatient, based on the traders' risk preference and strategies (Foucault et al. 2005). An alternate perspective to the traders' type, which is more relevant to price of underlying assets, classifies traders into fundamental and technical traders.

According to the efficient market hypothesis, the current price of one underlying asset reflects all information of past prices at least (Fama 1970). Fundamentalists tend to consider all information of their investment to decide their trading strategies. Since, good or bad news randomly happens and causes fluctuations in the price, fundamentalists generally trust that the abnormal price trend will go back to normal, and so, they take a long-term period in their trading.

Technical traders, however, are obsessed with past price chart. They believe market price trends can be repeated. If the frequency and speed is quick enough, they can follow price trend which is driven by someone else who can gain news of underlying assets earlierly, and benefit from high-frequency trading (Foucault et al. 2016). So that they use a series of trading rules based on past prices to make their high-frequency trading decisions. In other words, the motivation of past price trends indicates the possible change of current prices. Thus technical traders trust that they are able to capture the price changes earlier according to their designed rules, and then, they could make profitable strategies (Gencay 1999; Gencay and Stengos 1997).

With development of programming trading, more and more technical traders mix several technical trading rules in their strategies with the aid of computational power. They pursue higher speed and frequency to arbitrage so that they mostly prefer active and aggressive rules to construct orders in trading (Carrión 2013; Brogaard et al. 2014b). The original day-to-day trading strategies have evolved from minute-to-minute and even second-to-second. Hence, technical traders usually take short-term trading period and sometimes provide huge liquidity as “noise traders” in financial market (Tian et al. 2002). Technical analysis in that noise trader is quite simple. They can be recognised as price trend followers as well, which means that they buy when the price goes up and sell when the price goes down (or contrary to trend). This performance would bring higher profit than fundamentalists in the short-term, and also can be existing in long-term (De Long et al. 1990, 1991; Slezak 2003). Thus, how do technical traders lose or win the game in financial market through technical trading strategies? The aim of this paper is to investigate this question and try to disclose the traders' strategies and behaviours from the empirical points of view. Eventually, the work employs a set of applied economic and mathematical methods to investigate and explain technical traders and their strategies.

The structure of this paper is as below: Sect. 2 displays related literature and relevant source of method in this paper; Sect. 3 describes the details of the datasets in this empirical work; Sect. 4 details the process of filtering technical traders, and generation and matching of dummy trading signals; Sect. 5 shows works of traders' classification

by using k-means clustering algorithm and its results; Sect. 6 introduces a simple regression model to disclose the features of strategies in each group of traders; and Sect. 7 displays the conclusion and further works in the future of this paper.

## 2 Literature Review

This research concentrates mainly on the commodity futures, especially using rebar futures contract in Chinese futures market as the underlying asset to investigate technical trading strategies. Commodity futures market has many merits to discuss market participants' behaviours. It is well known that the main function of futures is to hedge investor's portfolio in order to guard against unexpected inflation or deflation in the future (Bodie and Rosansky 1980; Bodie 1983). In addition, one contract is similar as one share in stock market, which also has higher liquidity and lower cost to trade (Wang and Yu 2004). Compared with stock and other conventional financial market, commodity futures supply diversification and benefits for investment portfolios (Vrugt et al. 2004). Gorton and Rouwenhorst (2006) claim that commodity futures can afford the weak performance of stocks, due to unexpected inflation in a period. Erb and Harvey (2006) also support this point and suggest that active management of commodity futures can bring outstanding performance to investors' portfolios. Thus, various strategies of commodity futures have been discussed.

However, most of them pay attention to analysing applicability according to the profitability of different strategies. Vrugt et al. (2004), for example, proves that monetary policy and other related factors can construct profitable strategies in different commodity futures. It explains that fundamentalists would find it profitable based on external information of the underlying assets in a long-term investment, but this is not a standard method to evaluate the information effect on different portfolios. Regarding technical trading strategies, many empirical works have also indicated technical trading strategies can be profitable, and several of them believe that positive profits can be generated through different technical trading rules (Park and Irwin 2004). Donchian (1960) firstly states the channel trading rules in copper futures contract, and the following developed research on his work finds that the profitability of channel rules can exceed estimated transaction costs. For instances, 5.1–26.6% profit rate was generated by channel trading rules system in soybean, soybean meal, and soybean oil between 1984 and 1988 (Irwin et al. 1997). 3.8–5.6% mean returns are achieved by moving average and trade range break out systems in 12 futures contract, which include agricultural and metal commodity futures between 1978 and 1984 (Lukac et al. 1988). Lo et al. (2000), Neely (2002) and Faber (2007) find same evidence as Lukac's work in various financial markets. Jegadeesh and Titman (1993) raised momentum strategies, and Miffre and Rallis (2007) applied it to commodity futures, which brought the annual profit rate over 9%. Conversely, contrarian strategies generate abnormal returns in short-term investment of commodity futures (Lo and MacKinlay 1990). Cornell and Dietrich (1978) also document profitability of moving average and filter rules system with using Bretton Wood data. Brock et al. (1992) combines momentum-based moving average and trading range break out rules to investigate the performance of technical trading strategies. Certainly, the investigation of technical trading rules

has many valuable examples in foreign exchange market. Sweeney (1986) indicates that filter rules, which apply in ten kinds of currencies, can bring about 80% profits in the trading. Levich and Thomas (1993) find the application of filter and moving average rules is significant to trading profit of five currency futures. Longer moving average strategies can generate persisting profitability in emerging markets, which is founded by Fifield et al. (2008). Some related studies are Qi and Wu (2006), Sullivan et al. (1999), Neely and Weller (2003), Neely et al. (2009), Lucke (2003) and Marshall et al. (2008).

The above literatures discussed and invested the profitability and capacity of technical trading strategies in different financial markets. However, this paper tries to disclose whether technical participants employ technical trading strategies, what kind of technical trading rules would be used in their trading decisions, and how similar between traders' real actions and dummy actions according to different rules. Also, some previous surveys articles show the effect of technical trading strategies on individual trading behaviours, such as Lui and Mole (1998) and Oberlechner (2001). They find that market participants would adopt technical trading strategies in a lot for shorter forecasting intervals. The speed gain and act on price or news drives a "technological arms race" for technical traders in order to avoid missing out on potential benefit opportunities (Biais et al. 2015). According to Gehrig and Menkhoff (2006), the realisations of dummy buy or sell signals of different technical trading rules are generated by past price records. Then, market participants use and switch high-noise technical trading rules and low-noise fundamental trading rules to generate reasonable dynamics, volatility, uncorrelated profits and bubbles to carry out their trading strategies (Franke and Westerhoff 2012), even execution risk and cost would increase from crowding effect of technicalist with using arbitrage strategies in high frequency trading (Kozhan and Tham 2012; Brogaard et al. 2014a).

In this paper, we collect two unique datasets to make this relevant empirical works. The first key dataset is individual tick-by-tick data. The outstanding point of this data includes traders' identification and records all of their transaction details. Based on this element, we disclose technical traders' different strategies by real market participants. More details of this dataset descriptions are in Sect. 3. Another dataset is market price tick-by-tick data. It is utilised to generate dummy signals. It includes all price records also as tick-by-tick, high frequency, and per second data—which covers all selected underlying rebar futures contracts in the individual transaction dataset. After smoothing this dataset, we generate the dummy trading signals of selected technical trading rules via using smoothing per second price data.

For compiling technical trading strategies, we select three popular kinds of technical trading rules—Momentum, Moving Average, and Trading Range Break-out. According to intraday trading time in Chinese futures market, each kind of rules has different parameter settings and generates a technical strategies universe with 13,500 different rules. And, we introduce the contrarian rules of the above three rules. Hence, the amount selected rules in this research are  $13,500 * (3 + 3) = 81,000$ . Then, we combine the dummy signals and traders' real action together to calculate the similarity between them. Based on the similarities between each trader's real actions and each rules' dummy signals, we employ k-means clustering algorithm to classify technical traders. Referring to the clustering results, technical traders' strategies can be divided

into 11 groups in my dataset. At the end, on the basis of coordinates of 11 clusters, the characteristics of technical strategies in each group are disclosed.

### 3 Data Description

In the financial market, futures market is regarded as being a high-liquidity market, which is similar to stock and foreign currency market. Normally, most products on futures market are commodities, such as corn and copper. Previous research in this area mostly analysed stock index futures, however, commodity futures could be more influenced by macroeconomic effects and the risk level is higher than the other financial market due to its attribute of hedging and higher leverage rate (Fabozzi et al. 2008). Investors should be more careful about their investment than stock market due to the trading assets, which may reflect on the benefit of their business. Thus, if the fundamental traders occupied a lot of part of participants, it brings an advantage to this research that the pure technical traders could be captured by some special methods. If most main macroeconomic and other external elements are not able to influence traders' behaviour, the traders would be recognised as pure technical traders. Thus, this paper utilises the comprehensive tick-by-tick transaction data which can partly reflect market participants' trading behaviour, even better than stock and currency market to disclose technical traders. The trading behaviour can be more efficient and significant to capture, investigate, and identify.

The data collection in this paper is from one of most influential futures brokerage and one famous data statistics company in China. In the Chinese financial market, the role of market maker does not exist currently. Thus, the further complex consideration of market maker should be avoided. All market participants are under a fair market mechanism. The function of futures brokers is just to transfer the participants' order to the main exchange: Shanghai Futures Exchange, Zhengzhou Commodity Exchange, Dalian Commodity Exchange, and China Financial Futures Exchange. The data only covers one main commodity—Rebar (RB), which is used to trade at the Shanghai Futures Exchange. Even the Chinese futures market is unique to other main futures market, such as LME and CBO. This market also is one of the most active markets in the world. Nevertheless, Chinese financial derivative has even just started developing in the past 20 years, and financial derivatives had a speedy promotion until now. Futures is a relative measure product in Chinese financial derivative market and it plays an important role in the global futures market: e.g., Shanghai Futures Exchange is the second biggest exchange for copper trading in the world.

Specially, we choose rebar futures contract as the underlying asset by some following reasons. Firstly, the data is very hard to acquire and it is very precious for academic works because it generally records a fully real transaction order book (without entrust part) of one underlying asset rather than simulation. That is acquired during my intern period and accepted by my company for only using to academic works without any commercial purposes. Since we finish all relevant works, the data will be destroyed at the end. Thus, we hope to choose one asset which could reflect and represent Chinese futures market. Rebar futures would be one of the best relevant choice due to there is no other global futures markets launching this product, and Chinese rebar futures is

the most traded metal futures (by volumes and trading amount) currently in the whole world. That is the main reason that we choose rebar futures, rather than the well-known stock index futures or other commodity futures, as the unique dataset in this paper. Unfortunately, we only have the transaction data rather than a complete order book. The other entrust orders are very difficult to acquire.

Secondly, comparing with stock index futures, the margin requirement of rebar futures is quite low, in other words, rebar futures market is easier to entry than stock index futures in China. For instance, The lowest margin requirement for one unit of trading volume is equal to (volume \* leverage multiplier \* price (or index) \* margin ratio). The leverage multipliers of HuShen 300 futures (main index futures in China) and rebar futures are 300 and 10. That interprets the lowest margin requirement of HuShen 300 futures is 30 times higher than rebar futures. If the price and margin ratio are 3000 and 10% for both HuShen 300 futures and rebar futures, a trader only needs to pay  $1 * 10 * 3000 * 10\% = 3000$  RMB (local currency) to enter rebar futures market, but he has to pay  $1 * 300 * 3000 * 10\% = 90,000$  RMB to enter HuShen 300 futures market. Return to the aim of this paper, the number of market participants of rebar futures is absolutely greater than stock index futures, and lower margin requirement brings convenience to technical trading in rebar futures trading, which supports market participants are easy to form their technical trading strategies. Also, that indicates it is better to capture technical trading behaviours in rebar futures market.

Finally, technical trading, even programming trading is all-pervading in Chinese financial market. It interprets not only institutional traders utilize strict trading rules to execute trades but also individual traders adopt as well. Considering the background of the rebar futures, it was launched on 27th March, 2009. As we know, the main function of rebar is for building and infrastructure construction.

For China—the biggest developing country, real estate, industry, public equipment and many other social constructions cannot be promoted without rebar. Also, the Chinese mainland has a great inventory of iron ore, which is the raw material of rebar. Thus, the demand and supply volume of rebar is absolutely enough to support a high liquidity trading market. To reduce the risk, steel industries and steel trading business, which need or produce rebar, do not only consider the spot market, they also invest in the futures market for hedging or arbitrage. This is also an interesting and special futures market, in that, it looks like a pure speculative financial market.

Meanwhile, Shanghai Futures Exchange does not encourage participants to delivery real commodities after execution day, which means that nearly all of market participants must close out all of their positions before the end of each contract. According to mechanism of margin account in Chinese futures market, rebar future contracts have relatively lower requirement of margin, which cause individual market participants easily enter this market and make the highest average daily trading volume over all commodity futures due to the lower “threshold” with lower leverage ratio of rebar futures contracts trading produce similar high liquidity as stock market. Therefore, the above reason display rebar futures market is a really good sample to cover any kind of market participants. Due to the market is very related to macro economy, it also is a suitable sample to filter fundamentalist and technologist.

The utilised data in this paper includes two main datasets and one created timing announcement data series. The first data base (data1) is the tick-by-tick high frequency-

data of rebar contracts transaction (transaction order book, individual data) from the above mentioned futures company. The research period is from the starting date of rebar contract (27/03/2009) to the end of October in 2012 (31/10/2012). There are 19,933 traders, which include 19,760 individual and 173 institutional traders, taking part in rebar futures contract in this period. The data records cover each investor's transaction details, which includes contract code (identify different contracts), trader code (identify different traders which is the most outstanding point in this data), transaction time (accurate to seconds), transaction price, individual trading volume, individual position (net and real),<sup>1</sup> trading indication (open or offset and long or short),<sup>2</sup> and type-sign of investors (individual or institution). The total records of this database are 5,652,091.

We have to apologize that we only acquired transaction data without entrust data, so that the order book is complete to trader's transactions and incomplete to all actions of traders. It is hard and impossible to get full order book which includes all entrust orders. However, technical trading rules is designed and executed by procedure automatically. Once the market price or other required elements satisfy and trigger the given rules, the program will be as soon as possible to match market orders to execute trades according to the rules' decisions. Meanwhile, only selecting transaction orders is a right way to observe what kind of technical trading behaviours can be accepted or absorbed in the rebar futures market. Thus, we believe that is approving to only utilise transaction orders to observe trading behaviours.

Generally in futures market, different commodity futures contract may have different number of contracts in 1 year and also have different design of trading length (1 year, 6 months, etc.). For rebar futures contract, there are 12 contracts in 1 year. Each contract starts trading at the beginning of each month and delivery or execute at the same time (15th in each month) in the next year. For example, rb201006 started trading on June 16th 2009 and delivered on June 15th 2010. Because rebar contract launched at March, 2009, the first contract is named rb200909. Thus, the investigated data covers and indicates rebar futures contracts from rb200909 to rb201310 with contract code and the data has complete records of rb200909 to rb201210 and incomplete records of rb201211 to rb201310. These total 50 contracts establish a different cross-section to different traders. Trader code marks different 19,933 investors. As above, this paper only discusses investors' transaction part, the records only contain investors' transaction orders and do not cover the investors' other bid and ask orders.

The second dataset is the tick-by-tick high frequency data (data2) of the whole market price records. We collect the data from the mentioned data statistic company.<sup>3</sup> This dataset is different from data1. It just displays the whole market dynamics of transaction but not include any individual transaction details. The data records all transactions of each trading day during the research period and also includes transac-

<sup>1</sup> Net position means investor will sell or buy how many contracts of commodity on delivery day after one transaction.

<sup>2</sup> In the futures market, long means investors expect to buy futures contracts, and short means investors expect to sell futures contracts. Open and offset indicate the investors' real position. Investors can either take long or short position for their open position. Offset position is the opposite act to open position (Hull 2012).

<sup>3</sup> <http://www.gtafe.com>.

tion price, trading volume, and other information which can be matched with the first part of data. However, this paper proposes to investigate individual trading behaviour. This data does not attempt the identification of different investors. Thus, it is auxiliary data for the data1. The important role of the data2 is to provide total market position and generate dummy trading signals based on different technical trading rule in the following, which cannot be realised by data1.

Meanwhile, this paper also utilises some main macroeconomics index announcement to identify fundamental and technical traders which the method refers to Jiang et al. (2012). The rebar market is quite sensitive by government macroeconomic policy because of its main functions, as described above. Therefore, the third part of data is about the announcement time of the macroeconomic index. An examination of whether this public information can affect investors' preference and act could reveal potentially useful results.

We choose five key macroeconomic indexes which are quite related and influential to rebar market: Producer Price Index (PPI), Purchase Management Index (PMI), Real Estate Climate Index (REI), Entrepreneur Climate Index (EI), and Gross Domestic Product (GDP). The announcement time of PPI, PMI, and REI is announced monthly, and for EI and GDP is announced quarterly by the Chinese government. If the traders' trading volume is correlated with some indexes, it means that these traders are not purely technicalist. If trader's volumes do not have any relationships to these macroeconomic indices, these groups of traders would be recognised as pure technicalist.

## 4 Empirical Analysis

In the first step, we use a simple multiple regression model to filter pure technical traders in the data1. Before this work, there are two important problems requiring handling at first. One is endogenous of transaction price due to the data1 is just a part of the whole market and the other one is about removing irrational trading behaviour at the end of trading time of a single contract.

### 4.1 Handling Endogenous Variables and Irrational Transaction

The endogenous variable is the transaction price, which is unavoidable. In the regression model, we use the trading volume for each transaction to be the instrument of trader's dynamics and behaviour and other elements to be the explanatory variable for their behaviour, such as transaction price. It directly adopts transaction price rather than returns or profitability as an explanatory variable in the model. However, as discussed before, the owned transaction price data should be recognised as an endogenous variable since the data1 is just a part of the whole market. For instance, the data is just a part of the total market records. Other investors' behaviour (in the error term, cannot be observed) can not only influence market price change, but also impact the traders' trading volume in the researcher's owned data. On the one hand, outside of this owned data, investors affect market price changes—which also mean these acts impact inside sample investors' transaction prices. On the other hand, outside sam-

ple investors' trading volume can significantly affect inside sample traders' trading volume due to herding (Nofsinger and Sias 1999).

This paper refers to the theorem of Instrumental Variable Estimation (IV) to deal with this problem. Wooldridge (2011), if the instrumental variables are absolutely exogenous to the regression model, two-stage least square (2sls) can be dividedly achieved. Thus, we made four kinds of average price rely on the data1 and data2 as the instrumental variables to transaction price in this research sample: ap1—previously average 1 h total records' price of broker sample; ap2—previously average 500 records' price of broker sample; ap3—previously average 1 h total records' price of market sample; ap4—previously average 500 records' price of market sample. The ap1 and ap2 depend on the research sample, and ap3 and ap4 depend on the second part of total market data. These four IVs are the average price of the historical tick records. Thus, they are absolutely exogenous to the transaction price and trading volume. Also, all IVs have more than a 97% correlation with the transaction price in my original sample. Then, we adopt OLS to get the predict value of transaction price with the four IVs: ap1 to ap4 and the exogenous factors, individual position, and other factors. The price-hat (predict value of price) takes the place of original price in the regression model. Generally speaking, this method divides 2sls into two steps. It causes different stand errors for the final results. However, the significance will not have any changes.

This paper only pays attention on the significance of all the explanatory variables, which means it does not consider the coefficient. Thus, the dividing method is reasonable for utilization. Meanwhile, in the regression of this paper, the explained variable is trading volume of each transaction record. In order to reduce the simultaneity bias, we process the initial data. The initial data of trading volume is nominal—which contains the number of units of contract trade. Regarding the rebar futures contract, one trading unit of contract actually is equal to 10 tons of rebar, which is the leverage ratio of rebar futures contract. Thus, we adopt using nominal trading volume to multiply 10 to achieve real trading volume in order to deflate the price effect.

The second problem is called “irrational trading behaviour” which, firstly, occurs at the end of the trading time of a single futures contract. This behaviour means that investors' trading behaviour is uncorrelated with market price, holding position, and any other factors. Previous research on trading behaviour in futures market, generally speaking, missed this question. For one futures contract, it has its active period and also has its inactive period. For the instance of rebar futures, one contract change from active to inactive before three months to the execution date generally in China. But for the stock market, the trading time is continuous even if the listed company is delisted.

The question, then, is if an irrational trader still holds some contracts just before delivery day, he will drop his position even the market price is too unexpected for his portfolio. The reason is that such investors do not have real commodities for delivery, and also do not have utilization of commodities. The mechanism corresponds with the regulation of Shanghai Futures Exchange. Such market participants only have one aim is to speculate in this market. They just propose to speculate and not hedging or arbitrage. And, speculators occupy a huge part of futures markets so that most of them tend to clear out their position before execution date. Currently, there is no good method dealing with this problem. Thus, we require making a strong assumption in this paper that: The uncorrelated trading behaviour only occurs in the last two trading

month for each futures contract. This setting is based on the trend variation of total 50 contracts' market position and trading activity of the data1 and data2.

We checked all 50 rebar contracts' market position and trading activity trend. It is clear that all of contract's total market position and total trading volume has a significant decreasing trend in the last two months as we describe above according the the mechanism of SHFE. In other words, it means the contract becomes inactive generally two months before the delivery day. Therefore, we move out all the transaction records from data during last two months of each rebar contract. We also examine and use this new data and original data to do the same test. The results show that they are quite different that has many significance changes. Thus, we utilise this new data to continue the empirical research. The total records do not decline too much and just change to 3,893,880. This roughly method and weak consumption will be promoted in the future study.

## 4.2 Multiple Regression Model for Filtering Pure Technical Traders

In this section, we design a reasonable multiple regression model to filter pure technical traders for Chinese rebar futures. As we described before, we separate all participants into fundamentalist and technicalist. However, there must be a part of market participants mixing fundamental and technical knowledge to construct their trading decision. Thus, actually there are three types of traders: pure fundamentalist, “mixed”, and pure technicalist in the market. “mixed” traders will be also influenced by fundamental knowledge, such as news announcement so that we use the filtering model to also remove this group of traders, even they may adopt technical strategy in their trading decisions. Then, we delete any other “noisy” traders to only focus on pure technical traders and investigate their behaviours with concentration.

The filtering model is divided into two parts: the first part identify the fundamental relations between individual trading volume and market price (transaction price) and individual net position. The second part of data can indicate each investor's different trading behaviour, and show the difference of the above relations between each other and fundamental effects. The regression model is as below:

$$\begin{aligned} \ln(v) = & C + \alpha \ln(\text{price}) + \beta \ln(\text{position}) \\ & + \sum_{i=0}^n d_i [\alpha_i \ln(\text{price}) + \beta_i \ln(\text{position}) \\ & + \gamma_i \Delta T_{PPI} + \delta_i \Delta T_{PMI} + \varphi_i \Delta T_{REI} + \theta_i \Delta T_{EI} + \rho_i \Delta T_{GDP}] + \varepsilon \end{aligned}$$

where,

$$\text{position}_t = \text{market position}_t - \text{individual net position}_t$$

$$\Delta T_{macro-index} = \min(\Delta t, \Delta t + 1)$$

$$\Delta t + 1 = D_{A+1} - D_A$$

$$\Delta t = D_o - D_A$$

In the regression model,  $C$  is constant item,  $\varepsilon$  is error term,  $v$  is individual trading volume in each transaction, *price* is the above described price-hat instead of transaction price (market price) for each record, *position* is individual net position—which refers to the number of units of corresponding contracts each participant holds, implying the difference between opened and closed positions. In addition, we suppose to see the fluctuation of individuals' net position so that the *position* actually equal current total *market position* minus current *individual net position*. The total market position is invoked by market data and based on same time points in both transaction and market data. Then, the variable of position is the individual net position variation tendency.

We take the logarithm for these three variables in order to reduce the number size and decline the effect of heteroscedasticity. These first two items on the right hand can show the fundamental relationship between individual trading volume and two controlled factors (price and position).

For the following items,  $d$  is the dummy setting for different traders which depend on the size of research sample (can be set from 1 to 19,933 to identify different traders). In the bracket, the first two items indicate the difference of significance between the whole situation and each trader's situation, in other words, they implies trading responses of different investors.

The next group of variables describe announcement time of the above five macroeconomic variables. It is important to note that we do not use the real public value of macroeconomic news announcement. We only utilize the announcement date of each index to create announcement time-variation series. So,  $\Delta T_{macro-index}$  is the minimum value of  $\Delta t$  and  $\Delta t + 1$ , and they are the number of days between transaction occurred date and last announcement date(or between next announcement date and transaction occurred date). Assuming there is only one trading time line and existing  $D_A < D_o < D_{A+1}$ . Then,  $D_{A+1}$  is the next announcement date of each macroeconomic index,  $D_A$  is the last announcement date, and  $D_o$  is the transaction of each record occurred date.

The significance of each of these five time-variation series disclose whether the given trader's trading activity is affected by the dynamics of news announcement. Fundamentalist's trading enthusiasm should be influenced (or uninfluenced) with approaching (or away) to announcement date. Therefore, the setting of  $T$  is the time changing trend between monthly (or quarterly) announcement and next announcement time of each macroeconomic index. There is no collinearity between each others as the following Table 1.

**Table 1** Collinearity between each time variation variables of five macroeconomic Index

	tPMI	tPPI	tREI	tEI	tGDP
tPMI	1.0000				
tPPI	-0.0233	1.0000			
tREI	-0.0542	0.0037	1.0000		
tEI	-0.0801	0.0902	0.0709	1.0000	
tGDP	-0.0689	0.0867	0.1398	0.0106	1.0000

This performance is used to identify and disclose whether the trader may consider the macroeconomic information of these five indexes with the public time of indexes pass by. Pure technical traders ignore all other external elements and only focus on previous price, in other words, the relationship between this five variables and trading activity (trading volumes) is insignificant for each pure technicalist traders. The five relevant indices cannot influence decided trading volume of pure technical traders. We do not use the price of any other relevant commodities' price because that may cause the problem of endogenous as we described above. Relatively, Chinese scheduled news and macroeconomic index publications can influence fundamentalist behaviours more efficiently, thus we selected these news rather than related prices, such as the price of steel and iron ore.

In attention, the main function of this regression is based on the regression results of this five macroeconomic timing variables, which can indicate who are technical traders. Also, in order to identify whether investors tend to buy or sell, we split the data into long and short two groups.

The working sample is huge, so that the investigation is divided into two parts. The first part is working on total sample through all records. This initial research only chooses 100 investors (dummy setting:  $n = 100$ ) to analyses the individual trading behaviour in the bracket of the regression model. These 100 investors are the most active traders, who have the most transaction records, in my sample. No. 100 trader still has 3,726 records during the research period. And, the total records of top 100 traders are 863,953, which is about 22% of the total records. They seem to be using algorithm to execute their technical strategies at a high frequency level. Because they are the most active traders, they should have significance to investigate and summarise the total sample of technical traders.

In addition, these top 100 traders are organised according by the amount of their records. And, NO. 4, 8, 15, 70, 71, 77, 79 traders are institutional investors and others are individual investors. This status is also consistent with real situation that individual investors hold most amounts of market participants.

For the second part, it is the research on each single active futures contract. Since, rebar futures contract started on March 27th 2009, according to the situation of trading volume, market position, and trading amount, we find only September, October, November, and December contracts in 2009, and January, May, and October contract in 2010 to 2013 can be defined as relatively active contracts. This is caused by the seasonal economic cycle reason in China. Therefore, based on the variation trends of related elements, this part chooses the marked 15 contracts with boldface letter as the investigated single contract in the Table 2.

They are surveyed in the second part of each single futures contract. The method is same as the first part. But, the investigated active traders increased from 100 to 200 because the decline of sample size. These 200 investors are the top most trading people for each single rebar futures contract individually. Therefore, the total research samples are 16 (15 active contracts + 1 whole contract). Meanwhile, we have also made a secondary task. Based on individual position and immediate transaction price and their product (real transaction value), we assumed investors' wealth is the maximum product and sort the rank of them in each individual contract. That might help to identify how trader's endowments affect their trading behaviour.

**Table 2** Number of observations in each rebar futures contract

Contract code	Maximum market position	Total trading records	Contract code	Maximum market position	Total trading records
<b>rb0909</b>	<b>423,426</b>	<b>79,028</b>	<b>rb1110</b>	<b>888,390</b>	<b>363,451</b>
rb0910	397,842	30,407	rb1111	1534	119
<b>rb0911</b>	<b>895,870</b>	<b>142,224</b>	<b>rb1112</b>	584	2
<b>rb0912</b>	<b>1,008,870</b>	<b>190,001</b>	<b>rb1201</b>	<b>772,958</b>	<b>142,304</b>
<b>rb1001</b>	<b>1,140,896</b>	<b>243,295</b>	rb1202	2292	2
<b>rb1002</b>	<b>1,011,436</b>	<b>123,746</b>	rb1203	308	9
rb1003	195,814	19,258	rb1204	556	8
rb1004	53,498	1745	<b>rb1205</b>	<b>865,430</b>	<b>211,069</b>
<b>rb1005</b>	<b>1,074,692</b>	<b>374,967</b>	rb1206	1906	22
rb1006	53,270	1789	rb1207	0	0
rb1007	9626	335	rb1208	0	0
rb1008	17,078	1333	rb1209	2038	77
rb1009	26,948	1442	<b>rb1210</b>	<b>959,664</b>	<b>161,609</b>
<b>rb1010</b>	<b>1,497,516</b>	<b>619,286</b>	rb1211	1310	13
rb1011	26,292	1782	rb1212	1264	54
rb1012	36,544	1567	<b>rb1301</b>	<b>1,891,202</b>	<b>340,220</b>
<b>rb1101</b>	<b>1,426,002</b>	<b>417,135</b>	rb1302	3404	27
rb1102	21,718	427	rb1303	244	14
rb1103	13,024	658	rb1304	0	0
rb1104	13,044	358	<b>rb1305</b>	<b>1,114,852</b>	<b>95,830</b>
<b>rb1105</b>	<b>1,328,002</b>	<b>326,670</b>	rb1306	0	0
rb1106	1556	138	rb1307	0	0
rb1107	7690	114	rb1308	0	0
rb1108	506	4	rb1309	788	14
rb1109	2746	1289	rb1310	1790	38
Total					3,893,880

According to the regression results, if all of the five macroeconomic timing-variation variables are not significant to the traders' trading volume, which means the macro or external factors are not able to affect the trading behaviour, this group of traders should be recognised as pure technical traders. The initial results show how many technical traders of top 200 active traders existing in these 15 samples as the following Table 3.

After statistics, there are about 50 traders recognizing as pure technical traders in each contract, who both long and short do not have significance between macroeconomic indexes and their trading volume. Thus, we can say that the pure technical traders generally occupy about 25% of the top 200 most active traders and they are able to be recognized as pure technicalist to process further investigation.

**Table 3** Number of technical traders in long and short groups for 15 contracts

	Long	Short	Intersection		Long	Short	Intersection
rb0909	61	59	52	rb1105	69	65	52
rb0911	81	74	60	rb1110	77	69	57
rb0912	67	74	56	rb1201	70	61	52
rb1001	52	64	49	rb1205	59	68	48
rb1002	79	74	53	rb1210	55	78	46
rb1005	54	55	45	rb1301	65	62	50
rb1010	54	73	50	rb1305	79	105	56
rb1101	64	79	55				

#### 4.3 Investigate Sample Selecting and Selected Technical Trading Rules

The pure technical traders are selected by the filter model. However, some of them own fewer records in the sample of data1 (less observations). Thus, we select the research sample of traders who satisfy two conditions: 1, the traders must be pure technical traders who have been filtered. 2, the transaction records in their single trading contract must be more than 1500 records. Only few traders has amount of trades between 1000 and 1500 so that we set 1500 as the threshold to keep high frequency traders, who are more likely to use technical program trading. Therefore, we select traders from top 200 most active traders in each main futures contract. Some of them appear and can be selected in different contract, but we only choose one to symbolise this special traders. For instance, if trader 666 is identified as technical trader in two contracts, we only choose one contract as his research sample. After statistics, we choose 81 traders from each of 15 main contracts into the research sample. They are pure technical traders and have transaction records between 1500 and 12000. These 81 traders' behaviour can be representative to all of technical traders. All the following research is based on these 81 traders. Certainly, Technical Traders only focus on the historical price chart. They use the historical data to design a lot of different technical trading rules in order to execute their trading strategies. We select three kinds of popular technical trading rules as the bench mark of pure technical traders to investigate their behaviours.

Regard technical trading strategies, this research only selects three popular classes of technical trading rules (Momentum, Moving Average, and Trading Range Break-out). The signals of different rules are generated by the time division data. Based on the price trend and different regulations, the rules show the dummy trading signals at each time and the dummy signal may same as traders' real trading action if the traders follow the rule. This research also covers the contrarian rules of the three selected rules. The principles are same but the generated signals are opposite to the momentum. Thus, six kinds of rules are covered actually. The following descriptions include all details of each rule:

$P_t$ : market price of a future contract

$I_t$ : +1:long; -1:short; 0: keep neutral

$t \in \{1, 2, \dots, 13500\}$ : 1 trading day = 3.75 h = 225 min = 13,500 s

Trading time from SHANGHAI Futures Exchange in one day:

09:00–10:15: Trading (1.25 h) 10:15–10:30: Short Break

10:30–11:30: Trading (1 h)

11:30–13:30: Noon Break

13:30–15:00: Trading (1.5 h)

Total: 3.75 h

1. Momentum Rule (MO) refers to Conrad and Kaul (1998) and Chan et al. (2000). It is the basic rule of technical traders. The indicator shows whether market price change of a contract is positive or negative over a time period. If the current price is higher (or lower or equal) than the price at a defined time point, the rule would show the buy (or sell or keep nature) signals. The principle is that technical traders trust the price movements will bring the same price movements as before. It depends on the difference between current and previous price.

$$I_t(n) = \begin{cases} +1 & \text{if } P_t > P_{t-n} \\ 0 & \text{if } P_t = P_{t-n} \\ -1 & \text{if } P_t < P_{t-n} \end{cases}$$

Also, based on the momentum rule, we introduce Contrarian Rule (IMO) which is the opposite rule to momentum. They have same principle but inverse execution: when the price change is positive (or negative), the traders will sell (or buy).

$$I_t(n) = \begin{cases} +1 & \text{if } P_t < P_{t-n} \\ 0 & \text{if } P_t = P_{t-n} \\ -1 & \text{if } P_t > P_{t-n} \end{cases}$$

2. Moving Average Rule (MA) considers the weighting of all prices during a previously defined trading period. Most previous research covers this rule, such as Boswijk et al. (2001). Through calculating average price over a specific period, trader can identify whether traders act transaction. If the current price is higher (or lower or equal) than the average price during the previous trading period, the rule indicate the buy (or sell or keep nature) signals. We also introduce Contrarian Moving Average Rule (IMA), which have the inverse signals to MA.

$$MA : I_t(n) = \begin{cases} +1 & \text{if } P_t > MA_t \\ 0 & \text{if } P_t = MA_t \\ -1 & \text{if } P_t < MA_t \end{cases} \quad \& \quad IMA : I_t(n) = \begin{cases} +1 & \text{if } P_t < MA_t \\ 0 & \text{if } P_t = MA_t \\ -1 & \text{if } P_t > MA_t \end{cases}$$

Here,  $MA_t = \frac{\sum_{i=t-n+1}^t P_i}{n}$

3. Trading Range Breakout Rule (BO) is generally known as price channel systems. We refer to the literature from Park and Irwin (2004). When we define a specific trading period, BO shows a buy signal if the last (current) price is the highest price and generates a sell signal if the last (current) price is the lowest price during the

period. As the mention from Jackson and Ladley (2016), the principle of BO is to utilise the local maximum and minimum price as the motivation of technical traders in order to implement their strategies. Based on different periods, the trader seeks the extreme price as their “support”, and they trust that the price trend will follow this “support”. We still introduce the Contrarian Trading Range Breakout Rule (IBO) as the same principle of CMO and CMA.

$$BO : I_t(n) = \begin{cases} +1 & \text{if } P_t > \max(P_{t-1}, P_{t-2}, \dots, P_{t-n}) \\ -1 & \text{if } P_t < \min(P_{t-1}, P_{t-2}, \dots, P_{t-n}) \\ I_{t-1} & \text{otherwise} \end{cases}$$

$$IBO : I_t(n) = \begin{cases} +1 & \text{if } P_t < \max(P_{t-1}, P_{t-2}, \dots, P_{t-n}) \\ -1 & \text{if } P_t > \min(P_{t-1}, P_{t-2}, \dots, P_{t-n}) \\ I_{t-1} & \text{otherwise} \end{cases}$$

Before the first breakout, the indicator always stays equal to 0. After the first breakout, if the price does not satisfy the condition of changing indicator, the indicator follows the last (previous) indicator.

#### 4.4 Smoothing Data and Generation of Dummy Signals

All the above technical trading rules need to be calculated and generated by time division data. Most previous research used the data with same time interval, such as daily data. In other words, the data does not need to be modified (smoothing) because the same time interval is a kind of time series data and also it is the main feature of time division data. However, this paper utilises tick-by-tick data (both data1 and data2), which has the different time intervals for each record. Tick-by-tick data is the records of all transactions in the market. When one transaction happens, the data will add one record. Therefore, tick-by-tick data cannot be directly adopted to generate dummy signals of technical rules.

Even so, we use a general smoothing technique to transfer the tick-by-tick data to the time-series data in order to guarantee that there is only one price at each trading second. The aim is to use all the total market information to identify different trading rules. Thus, we utilise data2, which includes all ticks for all contracts, to smooth in order to generate dummy trading signals under different technical rules. We refer to use a simple method to fill empty record on the time series. This algorithm is described as below: we assume one of the occurred time point of record is “M” and the next time point is “N” (M and N are seconds). The prices of these two time points are  $P_M$  and  $P_N$ . Before time point “N”, the market price should keep the price of  $P_M$ , so that all time point should have a price record to complete smoothing aim and generate time division data:

$$\begin{aligned}
 P_M &= P_M \\
 P_{M+1} &= P_M \\
 P_{M+2} &= P_M \\
 &\vdots \\
 P_{N-2} &= P_M \\
 P_{N-1} &= P_M \\
 P_N &= P_N
 \end{aligned}$$

For instance of real process, the following Table 4 is the effective picture of data2 before and after smoothing algorithm.

The time series (time division) data has been created in the last step so that we start completing the generation of dummy trading signals with different rules. For the rules, we select 6 classes of rules, and each class of rules includes 13,500 (according to 13,500 s trading time in one trading day) types by different parameters. So, the amount of rules is  $6 * 13,500 = 81,000$  in the universe. Also, the main research contracts are 15 mentioned active contracts previously, thus we split 15 contracts as individual

**Table 4** Market tick-by-tick data before and after smoothing

Time	Price	Time	Price
2009/12/15 09:32:15	3770	2009/12/15 09:32:15	3770
2009/12/15 09:32:22	3769	2009/12/15 09:32:16	3770
2009/12/15 09:32:29	3771	2009/12/15 09:32:17	3770
2009/12/15 09:32:32	3769	2009/12/15 09:32:18	3770
2009/12/15 09:32:33	3769	2009/12/15 09:32:19	3770
2009/12/15 09:32:34	3768	2009/12/15 09:32:20	3770
2009/12/15 09:32:44	3769	2009/12/15 09:32:21	3770
2009/12/15 09:32:59	3770	2009/12/15 09:32:22	3769
2009/12/15 09:33:00	3769	2009/12/15 09:32:23	3769
2009/12/15 09:33:13	3767	2009/12/15 09:32:24	3769
2009/12/15 09:33:14	3769	2009/12/15 09:32:25	3769
2009/12/15 09:33:15	3768	2009/12/15 09:32:26	3769
2009/12/15 09:33:16	3769	2009/12/15 09:32:27	3769
2009/12/15 09:33:22	3770	2009/12/15 09:32:28	3769
:	:	2009/12/15 09:32:29	3771
		2009/12/15 09:32:30	3771
		2009/12/15 09:32:31	3771
		2009/12/15 09:32:32	3769
		2009/12/15 09:32:33	3769
		2009/12/15 09:32:34	3768
		2009/12/15 09:32:35	3768
		:	:

contracts to generate signals with 81,000 rules. Then, we make a huge technical data base (data3, it actually is “upgrade” and smoothing data of data2), which includes 15 files to indicate 15 contracts. In each file, it contains a matrix, where the column indicates 81,000 rules and the row indicates the price movement of the contract after smoothing data. Because the amount of observations of 15 contracts is not same, the size of matrix is not same. Thus, the columns (rules) are fixed as 81,000, and the rows are between 1,539,342 and 3,308,235.

Then, we have produced and introduced data3, which is very important to explain the effect of all selected technical rules. The next step is finding a way to link data3 to data1 in order to investigate pure technical traders’ strategies. It is difficult to investigate every trader’s specific behaviour, and so we tend to classify different types of technical traders. In each type, members should have generally similar strategies. In next section, we describe our method to link data1 and data3 and also show adopted classification method, which is K-means clustering algorithm.

## 5 Trader Classification

### 5.1 Data Reconstitution

There is a connection between data1 and data3. We adopt a simple and sensible method, which is to calculate the similarity (% , as percentage) between 81 traders’ real actions and dummy signals of each rule. In data3, as we introduced, it is time series data which means there is only one record (dummy signal) of one rule for all possible trading time. In data1, all traders’ real actions are included as –1 and 1 which indicate sell and buy with accurate time points. As above, the selected research sample covers 81 pure technical traders, and they have different amounts of transactions (observations or actions). We firstly filter 81 traders’ data from data1 to create 81 individual datasets. Next, we insert dummy signals of all rules from data3 into each individual dataset with considering same time points. Due to the range of time points of data3 contains all time points of data1, each trader’s real action must have their correspondently dummy signals with different rules with matching same time points. In other words, the originally individual dataset only include two columns—occurred time points and real actions, and now, 81,000 columns are added in the reconstituted dataset. Each column indicates the dummy trading signals of one specific technical rule. Then, we get 81 matrices for 81 traders, and the size of each matrix is: observations (row: each trader’s amount of records) multiply 81,002 (columns: 2 original columns include time point and real action, and 81,000 columns of dummy signals).

For each individual dataset, we can utilise basic computer techniques to calculate the similarity between traders’ real actions and dummy signals. This similarity implies what is the percentage of real actions same as each rule’s dummy signals. Thus, each of 81 matrices just provides one notice—the similarity. After statistic and combine all information from 81 matrices in one sheet, we achieve a significant matrix, and the size is  $81,000 * 81$  (include title row and first column, in fact, it is  $81,000 * 81$ ), which covers 81 traders’ similarity to all 81,000 rules. We put a short sample as below.

**Table 5** Similarity matrix between 81 traders and 81,000 technical trading rules

Rule	p01 (%)	p02 (%)	p03 (%)	p04 (%)	p05 (%)	p06 (%)
MO_13493	49.69	46.58	50.00	44.00	52.80	46.98
MO_13494	49.69	46.58	50.77	44.00	52.80	46.61
MO_13495	49.91	46.58	50.77	44.00	52.80	47.51
MO_13496	49.82	46.58	50.35	44.00	52.80	47.46
MO_13497	50.00	46.58	50.63	44.00	51.80	47.57
MO_13498	50.13	46.58	50.21	44.00	52.80	47.57
MO_13499	50.22	46.58	50.21	44.00	51.80	47.51
MO_13500	50.40	46.58	50.14	44.00	51.80	47.57
MA_1	19.66	23.29	35.44	31.20	36.10	25.92
MA_2	29.60	28.37	51.62	47.42	44.80	46.55
MA_3	36.50	34.01	58.65	54.07	50.70	55.85
MA_4	42.13	37.71	63.85	58.92	60.50	61.04
MA_5	46.00	41.68	67.23	58.48	62.80	61.46
MA_6	48.99	43.62	69.69	59.11	69.50	62.32
MA_7	51.50	45.10	68.00	60.78	69.40	63.76
MA_8	52.42	46.86	69.06	63.27	71.30	65.85
MA_9	53.39	47.50	71.80	63.46	71.50	67.45

In the Table 5, the first row indicates 81 traders from p01 to p81, and the first column indicates all 81,000 rules (MO\_1 to 13,500, MA\_1 to 13,500, BO\_1 to 13,500, IMO\_1 to 13,500, IMA\_1 to 13,500, IBO\_1 to 13,500). This matrix is the parent sample of traders' classification, because the similarities traders' different preference of each selected technical rules. Thus, we classify 81 traders in different groups based on the above mentioned similarity.

## 5.2 K-Means Clustering Algorithm

Cluster analysis<sup>4</sup> is used to classify many objects in different groups (clusters) with same features. In each group, there is a centroid, and all members have similar characteristics or coordinates to the centroid. Thus, we tend to adopt this method to group technical traders. There are various clustering algorithm. In statistical analysis, clustering analysis generally put all observations in a multi-dimensional space, and each observation becomes a point with n-dimensional attributes in the space (if the space has n dimensions). Based on the distance between each point and centroid, the algorithm select nearby points to each centroid as a group, which is centroid-based clustering. In my research, the similarities of each trader to each rule are seen as attributes in the

<sup>4</sup> For the Information and knowledge of clustering analysis, we refer to "Maimon and Rokach (2010), Data Mining and Knowledge Discovery Handbook. New York, Springer".

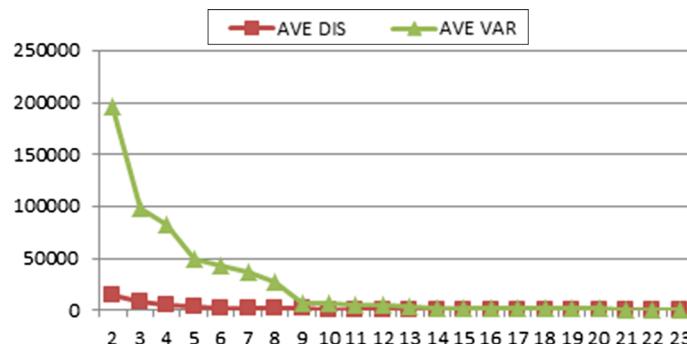
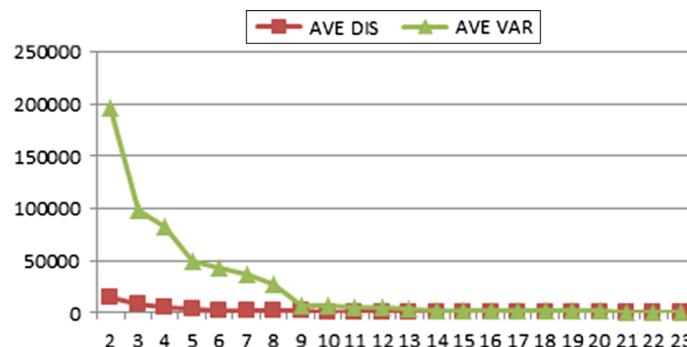
clustering space, so that clustering algorithm is easy and sensible to realise classification of technical traders. The logistic design is to put all 81 traders in the space: in other words, 81 points would be grouped. For every point, there are 81,000 attributes marking point's features, which imply the traders' different preferences for technical rules. Therefore, the above mentioned space is an 81,000-dimensional space for clustering. Centroid-based clustering generally has two popular ways. The first way is hierarchical clustering, and its principle is “from bottom to top”. Each point is one centroid at the beginning of clustering process. Then, the algorithm continuously merges close centroids to create a new centroid before finding the optimal number of centroid. After that, the process classifies all points in the space with optimal centroids. The other popular way of centroid-based clustering is k-means clustering, which we adopt in this work.

The main principle of k-means clustering is to partition all observations in the space into  $k$  groups. The clustering results also depend on the optimal distance, which is the least mean of all distance between member points and their individual centroids. The variety of distance can be appointed, such as city block and hamming distance. In this research, we utilise Matlab R2012b to realise k-means clustering because Matlab has standard procedure package of k-means. Also, we adopt the default distance—squared Euclidean distance (SED) of this automatic procedure. More details about the principle of SED are in “Appendix A”. The difference between k-means and hierarchical clustering is that we must appoint the number of  $k$  before analysis, for example, if  $k = 3$ , all points in the space will be divided into three groups. After identifying  $k$ , the algorithm starts stochastically set  $k$  centroids in the space. In the assignment step, k-means repeatedly moves the centroids until finding the optimal distance as above description. Then, the clustering is finished and we can get a sensible classification of traders.

There is one significant problem is that how to decide the number of  $k$ . We cannot randomly set a number of  $k$  with our “intuition”. There are many methods discussing in clustering area. In this paper, we refer to the method of “low average sum of variance and distance” (Maimon and Rokach 2010). The principle of this method is that, with increasing number of  $k$  setting, in each cluster, the average sum of distance between each point and their centroids and average sum of variance of distance in each group will be decreasing. When these two sums are close to 0 or at a lowest level in the dimension, they will not have a big change. Then, the corresponding number of  $k$  should be the decided and optimal  $k$  in the algorithm. Although, this is a roughly estimated method for  $k$  clusters, we designed three projects (three samples) to prove the correct number of  $k$ .

### 5.3 Clustering Results

In project one, we put the  $81,000 * 81$  matrix in the k-means algorithm and get the following statistic results, Fig. 1. Where, the x-axis is the number of  $k$ . We make 23 times of clustering with setting  $k$  equal 2 to 23. The y-axis is the value of average

**Fig. 1** ASD and ASV in project 1**Fig. 2** ASD and ASV in project 2

sum of distance and variance.<sup>5</sup> According to the graph, there is a huge decrease of ASD and ASV with increasing number of cluster (k). After k = 11, the ASD and ASV become stable. Thus in project one, all 81 traders should be divided into 11 groups with 81,000 attributes (rules).

**Project 1** It initially proves the number of classifications in the research, and it covers all investigated rules. However, this is a biased estimation, and the dimensional-space is very complex. Thus, we designed other two projects to support the clustering results. The principle is that we reduce the dimension of attributes in the space—We remove a lot of rules from the original 81,000 rules with two different criterions.

**Project 2** Setting intervals of the rules and choosing the rules to cluster, and the interval setting rule is in “Appendix B” ( $286 \times 6 = 1716$  rules). The feature here lies in choosing a rule every 5, 10, 15, 30 s before the 5 min ends, and then choosing a rule every 50 s until one day ends (13,500 s). We can see the variation in Fig. 2.

<sup>5</sup> Average sum of distance (ASD): After clustering, algorithm captures the sum of distance between every point and their attributive cluster (P to C distance) in k groups, and then gets average of k sums. Average sum of variance (ASV): Also, we write a procedure to calculate variance of P to C distance in every group. Less variance implies more stable and optimal members in each group.

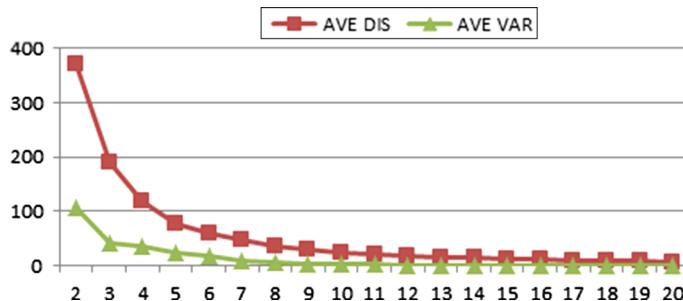


Fig. 3 ASD and ASV in project 3

**Project 3** Similar to P2, but it refers to the previous research. It has fewer rules to cluster, and the interval setting rule is in “Appendix C” ( $90 * 6 = 540$  rules). The feature is that choosing a rule every 5, 10, 15, 30, 60 (1 min), 600 (10 min), 1200 (20 min) seconds with increasing time. The Fig. 3 display the variation of ASD and ASV in project 3.

Then, the space becomes smaller (1716 and 540 dimensional-space) but the common features have not been changed because the selected rules are constructed by a standard and regular interval. It looks like the general sampling estimation. If the clustering results of this two small space are same as the original one, 11 should be recognised as the correct number of  $k$ . The above two graphs show the clustering results of project two and three. X-axis and y-axis have the same explanations of P1. It is important to note that it looks like that 9 or 10 would be the clustering results due to the dimension setting. However, there is still a length to 0 when cluster is 9 or 10. When cluster adopting 11, the length is really close to 0 and able to be acceptable.

We make 20 operations of P2 for the number of  $k$  equals 2–20 and 15 operations of P3 for the number of clusters equals 2–15. Also, due to the k-means algorithm randomly set the centroids, we run the program ten times for each project in order to get the relatively optimal clusters and keep a low level of variance. In the graph, it is very clearly that after grouping 81 traders into 11 groups, the ASD and ASV become stationary, which is the best evidence to support the clustering results of project one. The specific clustering results of P1, P2, and P3 are in the following Tables 6, 7 and 8 individually, we use 1–81 to label 81 traders.

Fortunately, no matter which project, some members are always in one group, in other words, the clustering of project one is successful. Therefore, all selected 81 pure technical traders can be classified in 11 groups with 81,000 attributes (technical trading rules).

In order to robustness check the clustering results, we split 81 traders into two groups: trader 1–40 as the first group and trader 41–81 as the other group. Then, we operate k-means clustering algorithm to each group with 81,000 rules. If the clustering results is similar as P1: around 11 centroids, we can confirm the results in nature. Figures 4 and 5 display the precious testing results.

Figure 4 shows the first 40 traders clustering results. As the same principle above, we can confirm 11 (or around this number) should be the correct number of centroids

**Table 6** Clustering results of project 1

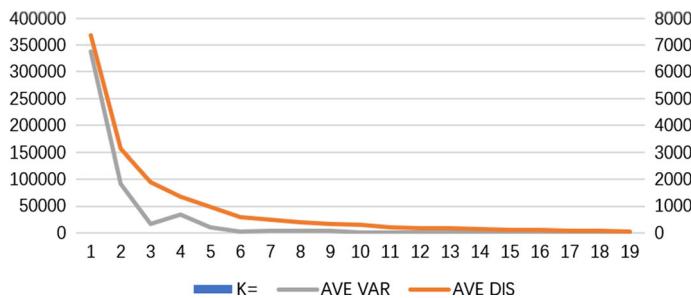
Group	Members in P1
1	1, 8, 9, 16, 20, 22, 24, 33, 34, 37, 38, 44, 46, 48, 50, 53, 62, 63, 74
2	15, 26, 57, 68, 71, 73
3	11, 21, 40, 41, 58, 75, 77, 79
4	3, 4, 12, 23, 27, 28, 36, 43, 51, 67, 69, 80
5	25, 39, 54, 66, 72
6	18
7	10, 59, 61
8	42, 47, 78
9	13, 14, 17, 19, 29, 30, 32, 52, 64, 81
10	2, 5, 6, 31, 55, 60, 65, 70, 76
11	7, 35, 45, 49, 56

**Table 7** Clustering results of project 2

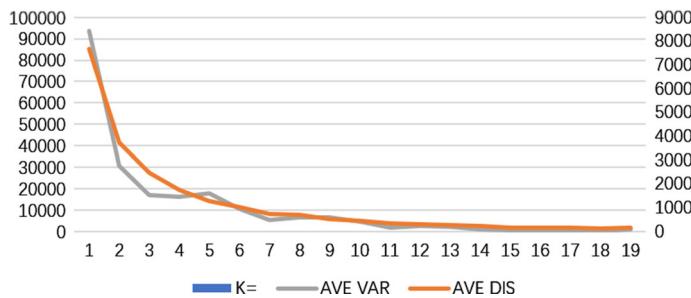
Group	Members in P2
1	1, 8, 16, 20, 22, 24, 31, 33, 34, 37, 38, 44, 46, 48, 50, 53, 60, 62, 63, 76
2	15, 26, 57, 68, 71, 73
3	54, 66, 67, 69, 72
4	10, 21, 40, 58, 75, 77, 79
5	4, 11, 12, 23, 25, 27, 36, 39, 41, 43, 51, 80
6	18
7	59, 61, 74
8	42, 47, 78
9	3, 6, 9, 13, 14, 17, 19, 28, 29, 30, 32, 52, 64, 81
10	2, 5, 55, 65, 70
11	7, 35, 45, 49, 56

**Table 8** Clustering results of project 3

Group	Members in P3
1	54, 66, 72
2	39, 41, 42, 51, 77
3	2, 5, 15, 26, 55, 57, 65, 70, 71, 73
4	18
5	4, 11, 75, 79
6	33, 47, 58, 62, 80
7	1, 3, 8, 20, 24, 27, 28, 34, 38, 50, 53, 61, 63, 67, 69
8	6, 9, 13, 14, 16, 17, 19, 29, 30, 32, 52, 64, 68, 74, 81
9	22, 31, 37, 44, 48, 59, 60, 76
10	10, 12, 21, 23, 36, 40, 43, 46, 78
11	7, 25, 35, 45, 49, 56



**Fig. 4** ASD and ASV in robustness check 1



**Fig. 5** ASD and ASV in robustness check 2

**Table 9** Clustering results of robustness check 1

Group	Members in robustness check 1
1	1, 8, 16, 20, 24, 33, 34, 37, 38
2	5, 15, 26
3	21
4	3, 4, 11, 12, 23, 27, 28, 36, 40
5	25, 39
6	18
7	10
8	14, 19
9	9, 13, 17, 29, 30, 32
10	2, 6, 22, 31
11	7, 35

in the space. It is also displayed in Fig. 5, we verify it again that these 81 pure technicalists' trading behaviours should be classified into 11 groups. For the members of each group, Tables 9 and 10 show the details. We can compare with the result in Project 1 (Table 6), the results are quite similar and further proof the clustering results are correct and suitable to this research.

**Table 10** Clustering results of robustness check 2

Group	Members in robustness check 2
1	41, 44, 46, 50, 51, 58, 62, 63, 64, 76, 80
2	73
3	75, 77, 79
4	66, 67, 69
5	49, 54, 72
6	74
7	59, 61
8	42, 47, 48, 53, 68, 78
9	43, 52, 81
10	55, 57, 60, 65, 70, 71
11	45, 56

## 6 Empirical Results—Trader's Characteristics

The key point of clustering is to use the centred coordinates to describe the members' attribute in each group. After clustering process of project one, 81 pure technical traders are divided into 11 groups. Thus, the 11 centroids summarise characteristics of members in their individual set. We operate the algorithm with setting  $k = 11$  about 50 times to achieve relatively least ASD and ASV. The results of coordinates construct an  $11 \times 81,000$  matrix, where 11 rows imply 11 centroids and 81,000 columns imply similarity degree with 81,000 technical rules. Using this matrix, we start further exploring trading strategies for each group. Based on the coordinates of centroid, which rule is much related to traders' real actions are very clear. If the similarity of the rule is higher, the rule is much more related to real actions. Then, the set of rules with higher similarity construct traders' strategies for each different group because the higher similarity implies the trader trend to adopt the rule. In the following steps, we select higher similarity rules and return to the dataset of dummy signals to draw out dummy signals according to the specific rules for every trader. Then, we make a simple regression between each trader's real action and dummy signals with rules of higher similarity, so that it supports the relationship between specific rules and traders' strategies. The selected technical rules cover six kinds of rules as mentioned before. Also, we extend the time length to 13,500 s, which is trading time in one day, for each variety. Thus, the amount rules are  $13,500 \times 6 = 81,000$  as discussed above. We adopt a simple regression model to investigate the effect of rules to traders' real actions as following:

$$S_j = C + \sum_{i=1}^{13500} MA_i + \sum_{i=1}^{13500} MO_i + \sum_{i=1}^{13500} BO_i \\ + \sum_{i=1}^{13500} IMA_i + \sum_{i=1}^{13500} IMO_i + \sum_{i=1}^{13500} IBO_i + \varepsilon$$

where,  $S$  is traders' real action as the dependent variable. The corner mark  $j$  of  $S$  is label of different traders from 1 to 81.  $MA$ ,  $MO$ ,  $BO$ ,  $IMA$ ,  $IMO$ ,  $IBO$  are corresponding signals with different rules to individual trader. In each kind of rule, it includes 13,500 rules. The constant term here is  $C$  and the error term is  $\varepsilon$ . The independent variables are all dummy signals with all 81,000 rules. Thus, we make 81 regressions with this model, and in each model, the total observations are equal to the total transaction records for each individual trader. However, this is the original investigated model. It cannot be realised due to the huge similar signals with different rules. In other words, the problem of collinearity happens. Each signals' variable only include 1, 0, and -1. The same or correlated variables possibly exist in total 81,000 rules (after experiment, same rules exist). The general methods of this question are to remove the same or correlated variables in the model. However, this is not advisable in this research. For example, if the vector of  $BO_{22}$  is equal to  $MA_{16}$  for trader 1, we do not know which rule we need to remove. If we remove both two rules, it influences the regression results because  $MA_{16}$  (or  $BO_{22}$ ) may be very significant to trader 1's real trading. So, we stop investigating all 81,000 rules and contact clustering results to seek some main significant rules with Top Six Project.

Top Six Project (T6) only selects the rules with highest similarity from  $MA$ ,  $MO$ ,  $BO$ ,  $IMA$ ,  $IMO$ , and  $IBO$ . After clustering, T6 chooses the most similarity rules from total six kinds of rules based on the centroid's coordinate in each group. Thus, 11 groups with top six highest rules are filtered. In each group, the six rules construct the key strategies of members. It is sensible to imply member's main strategies because the 11 centroid's coordinates indicate trading characteristics of all members in each group and also the top six rules' selection avoids "conflicts" between the same and correlated rules. Therefore, the original multiple-regression model can be transformed and simplified as:

$$S_{jg} = C + MA_{rg} + MO_{rg} + BO_{rg} + IMA_{rg} + IMO_{rg} + IBO_{rg} + \varepsilon$$

where,  $S$  still is traders' real action and the amount of observations is the total trader's individual transaction records. The following independent variables only content six specific rules based on the results of clustering. The corner mark  $g$  is from 1 to 11 which label the different group,  $j$  is from 1 to 81 which label the different traders, and  $r$  is the mark of rules which is selected from 1 to 13,500. As described before, based on the 11 clusters' coordinates, six rules of  $MA$ ,  $MO$ ,  $BO$ ,  $IMA$ ,  $IMO$ , and  $IBO$  with the greatest similarity can be selected in each group. They are components of T6's multiple regression. We return to the step of generating dummy trading signals of the six selected rules for each trader, and combine the six vectors and individual trader's real action in a new matrix. Then, we operate a T6's multiple regressions for each of 81 traders. The amounts of regressions are 81 and they actually divided into 11 groups as clustering. The regression results disclose the significance of selected rules to traders' real actions. The following Tables 11 and 12 summarise the final results.

Tables 11 and 12 explanations: The first row indicates the mean of each column. The first and second columns show the code of group and amount of members in the group. The third column shows the six specific rules with highest similarity in each group. Then, the following 4 columns show how many traders are affected by the

**Table 11** Significance percentage of top six technical trading rules in 1–7 clustering groups

Group	Amount of traders	Highest six rules	Significant traders	Significant probability (%)
1	19	bo13	17	89.47
		ma312	13	68.42
		mo251	12	63.16
		ibo990	10	52.63
		ima6355	9	47.37
		imo13406	8	42.11
2	6	bo351	5	83.33
		ma539	6	100.00
		mo387	6	100.00
		ibo3573	6	100.00
		ima9094	4	66.67
		imo6375	4	66.67
3	8	bo7320	3	37.50
		ma11723	4	50.00
		mo11973	5	62.50
		ibo122	7	87.50
		ima296	4	50.00
		imo235	6	75.00
4	12	bo1	10	83.33
		ma18	10	83.33
		mo13136	9	75.00
		ibo113	9	75.00
		ima11694	7	58.33
		imo7583	9	75.00
5	5	bo8	5	100.00
		ma41	5	100.00
		mo13114	5	100.00
		ibo453	5	100.00
		ima1136	3	60.00
		imo556	4	80.00
6	1	bo506	1	100.00
		ma3503	1	100.00
		mo1297	1	100.00
		ibo13476	0	0.00
		ima10	1	100.00
		imo9780	0	0.00

**Table 11** continued

Group	Amount of traders	Highest six rules	Significant traders	Significant probability (%)
7	3	bo13500	1	33.33
		ma13383	1	33.33
		mo12971	2	66.67
		ibo369	2	66.67
		ima1011	3	100.00
		imo887	2	66.67

**Table 12** Significance percentage of top six technical trading rules in 8–11 clustering groups

Group	Amount of traders	Highest six rules	Significant traders	Significant probability (%)
8	3	bo2484	1	33.33
		ma5232	2	66.67
		mo3071	2	66.67
		ibo94	3	100.00
		ima212	2	66.67
		imo275	2	66.67
9	10	bo32	6	60.00
		ma65	10	100.00
		mo92	8	80.00
		ibo5425	9	90.00
		ima13393	7	70.00
		imo11704	6	60.00
10	9	bo40	8	88.89
		ma215	8	88.89
		mo193	8	88.89
		ibo11156	6	66.67
		ima12324	4	44.44
		imo13146	7	77.78
11	5	bo13227	5	100.00
		ma20	4	80.00
		mo13414	2	40.00
		ibo68113	4	80.00
		ima2170	5	100.00
		imo536	5	100.00

six rules and how many are not, and the probability of total traders. For instance, the second big row actually includes 19 regressions for 19 traders in group 1. The rule of BO13 has effect to 17 traders in this group, and the occupation is 89.47%.

If we set 60% as standard, this group of traders tends to use bo13, ma312, and mo251 in their strategy. Traders in group 2 utilise ma539, mo387, and ibo3575 in their strategy absolutely. Still considering 60%, traders tend to put mo11973, imo235, and ibo122 in their strategy in group 3. In group 4, most traders adopt bo1 and ma18 rules. All traders utilise bo8, ma41, and mo13114 in their strategy. There is only one trader in this group. But, he is very interesting because he also is single trader in one group of project 2 and 3. The size of trader is small in this group so that the adopted rules are not clear. The rule of ibl94 is adopted for this group of traders. This also is a small group so that the indication is not very clear. These 10 traders utilise ma65 and tend to adopt ibo5425 and mo92 in their strategy. Most traders in this group adopt bo40, ma215, and mo193 in their strategy. The five traders in the last group use the ima2170, imo536, and bo13227 in their strategy. The results are according to the coordinates of clusters in each group. Hence, the strategies set can prefer the features of technical traders. Market participants can follow or “snipe” technical trading in the financial market based on the strategies in each group.

## 7 Conclusion

This is an empirical research on technical trading strategies. The contribution of this research is trying to indicate and disclose technical trading behaviour in Chinese rebar futures market and create a new method to capture technical trading strategies. We selected rebar futures contracts, which could be recognized as a representative commodity futures in Chinese futures market, as the underlying asset to investigate. According to the unique feature of the dataset, traders have their own identification. The top 100 most active traders were the main research object since they were more likely to be the technical traders and employ program trading. We chose five related macroeconomic indexes to rebar market as the filter factor by using a simple multiple-regression model to filter technical traders. The results showed that in each contract, technical traders occupied about 20% from the top 100 most active traders. Since it is high frequency per second data, we used similar tick-by-tick data which recorded the complete market price trend of each contract without traders’ identification, to generate dummy signals with a series of technical trading rules. Then, the dummy signals of each selected technical rules and each trader’s real action were combined and matched according to same transaction time. We calculated the similarity between them which indicated each trade’s motivation to employ each rule. We selected only 81 technical traders from 15 most active contracts in my dataset for investigation. Based on the similarity matrix, we adopted k-means clustering algorithm to classify these 81 traders. The clustering results showed that they could be divided into 11 groups with different technical strategies. In order to avoid same or correlated dummy signals with different rules, we chose the top six highest rules based on the coordinates of 11 clusters to state whether these six rules were significant to trader’s real actions. The results indicated that most members, in the different groups, had to have one or more significant technical rules to their real action. More details are displayed in Sect. 5. We will continuously improve our dataset in two part: cover all commodities in the market and bring more types of technical rules in the system. Then, that will be more significant

that we will test whether the summarized strategies are profitable in each group of technicalists. There would be some interesting things appearing in our further works.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendices

### A Principle of K-Means Clustering with Squared Euclidean Distance

We use  $p_{c,n,d}$  to indicate the coordinates of all points in the space and  $ce_{c,n,d}$  to indicate the centroids' coordinates. Where in the lower right corner,  $c$  is the mark of cluster (group),  $n$  is the number of points, and  $d$  is the dimension of the space. For instance, 4 traders are labelled as 1, 2, 3, 4 in an  $x$  dimensional-space, and their coordinates are “ $p_{c,1,x}, p_{c,2,x}, p_{c,3,x}, p_{c,4,x}$ ”. If setting  $k = 2$  in the algorithm, the procedure estimates two original centroids in the space, and their coordinates are “ $ce_{2,n,x}, ce_{2,n,x}$ ”. Also, the algorithm randomly estimates point 1 and 2 attribute to cluster 1 and point 3 and 4 attribute to cluster 2. Then, the algorithm starts doing the first calculation of SED between points and centroids.

Squared Euclidean Distance (SED):

$$d(P, Q) = (p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2 = \sum_{i=1}^n (p_i - q_i)^2$$

For centroid one to point 1 and 2:

$$\begin{aligned} d(ce_{1,n,x}, p_{c,1,x}) &= \sum_{i=1}^x (ce_{1,n,i} - p_{c,1,i})^2 \\ d(ce_{1,n,x}, p_{c,2,x}) &= \sum_{i=1}^x (ce_{1,n,i} - p_{c,2,i})^2 \end{aligned}$$

For centroid two to point 3 and 4:

$$\begin{aligned} d(ce_{2,n,x}, p_{c,3,x}) &= \sum_{i=1}^x (ce_{2,n,i} - p_{c,3,i})^2 \\ d(ce_{2,n,x}, p_{c,4,x}) &= \sum_{i=1}^x (ce_{2,n,i} - p_{c,4,i})^2 \end{aligned}$$

Then, the procedure calculates the average distance in each group:

$$\begin{aligned} \text{Average Distance of Group 1 (avdg1)} &= \frac{d(ce_{1,n,x}, p_{c,1,x}) + d(ce_{1,n,x}, p_{c,2,x})}{2} \\ \text{Average Distance of Group 2 (avdg2)} &= \frac{d(ce_{2,n,x}, p_{c,3,x}) + d(ce_{2,n,x}, p_{c,4,x})}{2} \end{aligned}$$

Based on  $avdg1$  and  $avdg2$ , two centroids start moving to a new position in order to get the optimal distance in each group. The centroids' movement will change the

members in each group and calculate new  $avdg1$  and  $avdg2$ . After a lot of time, a set of  $avdg1$  and a set of  $avdg2$  are created, and also the optimal clusters are achieved based on optimal average distance:

$$\text{Optimal Average Distance of Group1} = \min.(avdg1)$$

$$\text{Optimal Average Distance of Group2} = \min.(avdg2)$$

The finally optimal centroids are captured, and their coordinates is identified with x dimensions. These coordinates is very significant because it is recognised as implying the attributes of all members in each group. In this paper, 81 traders are divided into 11 groups with an 81,000 dimensional space. The coordinates are the similarity between traders' real actions and dummy actions with different technical trading rules. Thus, the clusters' coordinates indicate and imply the trading strategies of members in each group.

## B Interval Setting of K-Means Project 2

See Table 13.

**Table 13** Interval setting of k-means project 2

	Sec.	Min.	Hour	Day (3.75 h per day)
Start 5 s interval	1	0.017	0.000	0.000
	5	0.083	0.001	0.000
	10	0.167	0.003	0.001
	15	0.250	0.004	0.001
	20	0.333	0.006	0.001
	25	0.417	0.007	0.002
Start 10 s interval	30	0.500	0.008	0.002
	40	0.667	0.011	0.003
	50	0.833	0.014	0.004
Start 15 s interval	60	1.000	0.017	0.004
	75	1.250	0.021	0.006
	90	1.500	0.025	0.007
	105	1.750	0.029	0.008
Start 30 s interval	120	2.000	0.033	0.009
	150	2.500	0.042	0.011
	180	3.000	0.050	0.013
	200	3.333	0.056	0.015
	210	3.500	0.058	0.016
	240	4.000	0.067	0.018
	250	4.167	0.069	0.019
	270	4.500	0.075	0.020

**Table 13** continued

	Sec.	Min.	Hour	Day (3.75 h per day)
Start 50 s interval	300	5.000	0.083	0.022
	350	5.833	0.097	0.026
	400	6.667	0.111	0.030
	450	7.500	0.125	0.033
	500	8.333	0.139	0.037
	550	9.167	0.153	0.041
	600	10.000	0.167	0.044
	650	10.833	0.181	0.048
	700	11.667	0.194	0.052

## C Interval Setting of K-Means Project 3

See Table 14.

**Table 14** Interval setting of k-means project 3

	Sec.	Min.	Hour	Day (3.75 h per day)
Start 5 s interval	5	0.083	0.001	0.000
	10	0.167	0.003	0.001
	15	0.250	0.004	0.001
	20	0.333	0.006	0.001
	25	0.417	0.007	0.002
	30	0.500	0.008	0.002
	35	0.583	0.010	0.003
	40	0.667	0.011	0.003
	45	0.750	0.013	0.003
	50	0.833	0.014	0.004
	55	0.917	0.015	0.004
Start 10 s interval	60	1.000	0.017	0.004
	70	1.167	0.019	0.005
	80	1.333	0.022	0.006
	90	1.500	0.025	0.007
	100	1.667	0.028	0.007
	110	1.833	0.031	0.008
	120	2.000	0.033	0.009
	130	2.167	0.036	0.010
	140	2.333	0.039	0.010
	150	2.500	0.042	0.011

**Table 14** continued

	Sec.	Min.	Hour	Day (3.75 h per day)
	160	2.667	0.044	0.012
	170	2.833	0.047	0.013
	180	3.000	0.050	0.013
	190	3.167	0.053	0.014
	200	3.333	0.056	0.015
	210	3.500	0.058	0.016
	220	3.667	0.061	0.016
	230	3.833	0.064	0.017
	240	4.000	0.067	0.018
	250	4.167	0.069	0.019
	260	4.333	0.072	0.019
	270	4.500	0.075	0.020
	280	4.667	0.078	0.021
	290	4.833	0.081	0.021
Start 15 s interval	300	5.000	0.083	0.022
	315	5.250	0.088	0.023
	330	5.500	0.092	0.024
	345	5.750	0.096	0.026
	360	6.000	0.100	0.027
	375	6.250	0.104	0.028
	390	6.500	0.108	0.029
	405	6.750	0.113	0.030
	420	7.000	0.117	0.031
	435	7.250	0.121	0.032
	450	7.500	0.125	0.033
	465	7.750	0.129	0.034
	480	8.000	0.133	0.036
	495	8.250	0.138	0.037
	510	8.500	0.142	0.038
	525	8.750	0.146	0.039
	540	9.000	0.150	0.040
	555	9.250	0.154	0.041
	570	9.500	0.158	0.042
	585	9.750	0.163	0.043

**Table 14** continued

	Sec.	Min.	Hour	Day (3.75 h per day)
Start 30 s interval	600	10.000	0.167	0.044
	630	10.500	0.175	0.047
	660	11.000	0.183	0.049
	690	11.500	0.192	0.051
	720	12.000	0.200	0.053
	750	12.500	0.208	0.056
	780	13.000	0.217	0.058
	810	13.500	0.225	0.060
	840	14.000	0.233	0.062
Start 1 min interval	870	14.500	0.242	0.064
	900	15.000	0.250	0.067
	960	16.000	0.267	0.071
	1020	17.000	0.283	0.076
	1080	18.000	0.300	0.080
	1140	19.000	0.317	0.084
	1200	20.000	0.333	0.089
	1260	21.000	0.350	0.093
	1320	22.000	0.367	0.098
	1380	23.000	0.383	0.102
	1440	24.000	0.400	0.107
	1500	25.000	0.417	0.111
	1560	26.000	0.433	0.116
	1620	27.000	0.450	0.120
Start 10 min interval	1680	28.000	0.467	0.124
	1740	29.000	0.483	0.129
	1800	30.000	0.500	0.133
	2400	40.000	0.667	0.178
	3000	50.000	0.833	0.222
Start 30 min interval	3600	60.000	1.000	0.267
	4500	75.000	1.250	0.333
	6300	105.000	1.750	0.467
	8100	135.000	2.250	0.600
	9900	165.000	2.750	0.733
One day	11,700	195.000	3.250	0.867
	13,500	225.000	3.750	1.000

## References

- Bagehot, W. (1971). The only game in town. *Financial Analysts Journal*, 27(2), 12–14.
- Biais, B., Foucault, T., & Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial Economics*, 116(2), 292–313.
- Bodie, Z. (1983). Commodity futures as a hedge against inflation. *The Journal of Portfolio Management*, 9(3), 12–17.
- Bodie, Z., & Rosansky, V. I. (1980). Risk and return in commodity futures. *Financial Analysts Journal*, 36(3), 27–39.
- Boswijk, P., Griffioen, G., & Hommes, C. (2001). *Success and failure of technical trading strategies in the cocoa futures market*. Technical report, Tinbergen Institute Discussion paper 01-016/1, Amsterdam and Rotterdam University.
- Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5), 1731–1764.
- Brogaard, J., Hendershott, T., Hunt, S., & Ysus, C. (2014a). High-frequency trading and the execution costs of institutional investors. *Financial Review*, 49(2), 345–369.
- Brogaard, J., Hendershott, T., & Riordan, R. (2014b). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8), 2267–2306.
- Carrión, A. (2013). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, 16(4), 680–711.
- Chan, K., Hameed, A., & Tong, W. (2000). Profitability of momentum strategies in the international equity markets. *Journal of Financial and Quantitative Analysis*, 35(2), 153–172.
- Conrad, J., & Kaul, G. (1998). An anatomy of trading strategies. *Review of Financial Studies*, 11(3), 489–519.
- Cornell, W. B., & Dietrich, J. K. (1978). The efficiency of the market for foreign exchange under floating exchange rates. *The Review of Economics and Statistics*, 60(1), 111–120.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703–738.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1991). The survival of noise traders in financial markets. *The Journal of Business*, 64(1), 1–19.
- Donchian, R. D. (1960). Commodities: High finance in copper. *Financial Analysts Journal*, 16(6), 133–142.
- Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, 62(2), 69–97.
- Faber, M. T. (2007). A quantitative approach to tactical asset allocation. *The Journal of Wealth Management*, 9(4), 69–79.
- Fabozzi, F. J., Fuss, R., & Kaiser, D. G. (2008). *The handbook of commodity investing*. London: Wiley.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Fifield, S., Power, D., & Knipe, D. (2008). The performance of moving average rules in emerging stock markets. *Applied Financial Economics*, 18(19), 1515–1532.
- Foucault, T., Hombert, J., & Roșu, I. (2016). News trading and speed. *The Journal of Finance*, 71(1), 335–382.
- Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4), 1171–1217.
- Franke, R., & Westerhoff, F. (2012). Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. *Journal of Economic Dynamics and Control*, 36(8), 1193–1211.
- Gehrig, T., & Menkhoff, L. (2006). Extended evidence on the use of technical analysis in foreign exchange. *International Journal of Finance and Economics*, 11(4), 327–338.
- Gencay, R. (1999). Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, 47(1), 91–107.
- Gencay, R., & Stengos, T. (1997). Technical trading rules and the size of the risk premium in security returns. *Studies in Nonlinear Dynamics & Econometrics*, 2(2), 23–34.
- Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, 62(2), 47–68.
- Hull, J. C. (2012). *Options, futures, and other derivatives*. Essex: Pearson Education Limited and Associated Companies throughout the world.

- Irwin, S. H., Zulauf, C. R., Gerlow, M. E., & Tinker, J. N. (1997). A performance comparison of a technical trading system with arima models for soybean complex prices. *Advances in Investment Analysis and Portfolio Management*, 4, 193–203.
- Jackson, A., & Ladley, D. (2016). Market ecologies: The interaction and profitability of technical trading strategies. *International Review of Financial Analysis*, 47, 270–280.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91.
- Jiang, G. J., Lo, I., & Valente, G. (2012). *Highfrequency trading in the US treasury market: Evidence around macroeconomic news announcements*. Technical report, Working Paper, Washington State University.
- Kozhan, R., & Tham, W. W. (2012). Execution risk in high-frequency arbitrage. *Management Science*, 58(11), 2131–2149.
- Levich, R. M., & Thomas, L. R. (1993). The significance of technical trading-rule profits in the foreign exchange market: A bootstrap approach. *Journal of International Money and Finance*, 12(5), 451–474.
- Lo, A. W., & MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial studies*, 3(2), 175–205.
- Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, 55(4), 1705–1765.
- Lucke, B. (2003). Are technical trading rules profitable? Evidence for head-and-shoulder rules. *Applied Economics*, 35(1), 33–40.
- Lui, Y.-H., & Mole, D. (1998). The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence. *Journal of International Money and Finance*, 17(3), 535–545.
- Lukac, L. P., Brorsen, B. W., & Irwin, S. H. (1988). A test of futures market disequilibrium using twelve different technical trading systems. *Applied Economics*, 20(5), 623–639.
- Maimon, O., & Rokach, L. (2010). *Data mining and knowledge discovery handbook*. Berlin: Springer.
- Marshall, B. R., Cahan, R. H., & Cahan, J. M. (2008). Does intraday technical analysis in the us equity market have value? *Journal of Empirical Finance*, 15(2), 199–210.
- Miffre, J., & Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking and Finance*, 31(6), 1863–1886.
- Neely, C. J. (2002). The temporal pattern of trading rule returns and exchange rate intervention: Intervention does not generate technical trading profits. *Journal of International Economics*, 58(1), 211–232.
- Neely, C. J., & Weller, P. A. (2003). Intraday technical trading in the foreign exchange market. *Journal of International Money and Finance*, 22(2), 223–237.
- Neely, C. J., Weller, P. A., & Ulrich, J. M. (2009). The adaptive markets hypothesis: Evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis*, 44(2), 467–488.
- Nofsinger, J. R., & Sias, R. W. (1999). Herding and feedback trading by institutional and individual investors. *The Journal of Finance*, 54(6), 2263–2295.
- Oberlechner, T. (2001). Importance of technical and fundamental analysis in the european foreign exchange market. *International Journal of Finance and Economics*, 6(1), 81–93.
- Park, C.-H., & Irwin, S. H. (2004). *The profitability of technical analysis: A review*. Technical Report, AgMAS Project Research Report.
- Qi, M., & Wu, Y. (2006). Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market. *Journal of Money, Credit and Banking*, 38(8), 2135–2158.
- Slezak, S. L. (2003). On the impossibility of weak-form efficient markets. *Journal of Financial and Quantitative Analysis*, 38(3), 523–554.
- Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54(5), 1647–1691.
- Sweeney, R. J. (1986). Beating the foreign exchange market. *Journal of Finance*, 41(1), 163–182.
- Tian, G. G., Wan, G. H., & Guo, M. (2002). Market efficiency and the returns to simple technical trading rules: New evidence from us equity market and chinese equity markets. *Asia-Pacific Financial Markets*, 9(3), 241–258.
- Vrugt, E. B., Bauer, R., Molenaar, R., & Steenkamp, T. (2004). *Dynamic commodity timing strategies*. Technical report, working paper, Maastricht University.
- Wang, C., & Yu, M. (2004). Trading activity and price reversals in futures markets. *Journal of Banking and Finance*, 28(6), 1337–1361.
- Wooldridge, J. M. (2011). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.