Technical Trading Behavior: Evidence from Chinese Futures Market

Guanqing Liu

Department of Economics, University of Leicester

December 16, 2013

Abstract:

Technical Traders adopt mathematical methods to formulate various technical trading rules on their trading strategies. This paper utilizes 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. I firstly use a simple multiple regression model to filter technical traders in individual dataset. Then, dummy signals according to 6 popular kinds of technical rules are generated as benchmark by market dataset for real trading actions. According to similarity between dummy signals with different technical rules and traders' real actions, I employ k-means algorithm to classify technical traders. Through these empirical works, technical traders in my dataset are classified by 11 groups. On the basis of 11 clusters' coordinates, features of technical strategies in each group are summarized finally.

Keywords: rebar futures market, technical trading, k-means clustering

1: Introduction and Literature Review

Trader's types have many kinds of identification. Current theoretical researches on market microstructures usually divide traders into informed and uninformed traders (Bagehot & Treynor, 1971). With development, traders also are classified as patient and impatient based on traders' risk preference and strategies (Foucault, Kadan, and Kandel, 2005). Another perspective to the traders' types, which is more relevant to the price of underlying assets, traders can be grouped into fundamental and technical traders. As efficient market hypothesis, current price of one underlying asset reflects all information of past price at least (Fama, 1970). Fundamentalists tend to consider all information of their investment to decide their trading strategies. Because good or bad news randomly happens and cause the price away from the fundamental price, fundamentalist generally trust the abnormal price trend would back to normal price level and so that take long-term position in their trading. Oppositely, technical traders are obsessed with past price chart. They believe market price trend can be repeated and followed so that they use a series of trading rules based on past price to make their trading decisions. In other words, the motivation of past price trend indicate the possible change of current price, thus technical traders trust they are able to recognize the change earlier with different rules in order to make a profitable strategy (Gencay, 1999). With development of programming trading, more and more technical traders mix several of technical trading rules in their strategies with the aid of computational power. The original day-to-day trading strategies have evolved to minute-to-minute even second-to-second. Hence, technical traders usually take short-term position and sometimes provide huge liquidity as noise traders in financial market (Tian, Wan, and Guo, 2002). Technical analysis in that noise trader is quite simple. They can be recognized as price trend follower as well, which means buy when the price goes up and sell when the price goes down (or contrarian to trend). This performance would bring higher profit than fundamentalist in short-term and also can be existing in long-term (De Long, Shleifer, Summers, and Waldmann, 1990 &1991; Slezak, 2003). So, how technical traders make profit or loss and win the game in financial market through technical trading strategies? The aim of this paper is to investigate this question and try to disclose their strategies and behaviors from the empirical points of view. The work employs a set of applied economic and mathematical methods to investigate and explain technical traders and their strategies finally.

This research employs one main commodity futures: rebar 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 against unexpected inflation or deflation in the future (Bodie, 1980 & 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). Compare 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 commodity futures can offset the weak performance of stocks due to unexpected inflation in a period. Erb and Harvey (2006) also support this point and suggest 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 on analysing applicability according to profitability of different strategies. Vrugt et al. (2004) proofs monetary policy and other related factors can construct profitable strategies in different commodity futures. It explains fundamentalist would be profitable based on external information of the underlying assets in a long-term investment, but this is not a standard method to evaluate information effect on different portfolios. Regard technical trading strategies, many empirical works have also indicated technical trading strategies can be profitable and a lot of them support positive profits can be generated by different technical trading rules (Park and Irwin, 2007). Donchian (1960) firstly employs channel trading rules in copper futures contract, and the following developed research on his work examine the profitability of channel rules can exceed estimated transaction costs. For instances, 5.1%-26.6% profit rate is 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 market. Jegadeesh and Titman (1993) raised momentum strategies, and Miffre and Rallis (2007) apply it to commodity futures, which bring 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 (1987) also document profitability, but, of moving average and filter rules system with 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 good examples in foreign exchange market. Sweeney (1986) indicates 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). More related literatures are according to Qi and Wu (2006), Sullivan et al. (1999), Neely et al. (2003 and 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 market. However, this paper tries to disclose that which technical trading rules or what set of rules to market participants would really employ in their trading decisions. 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 market participants adopt technical trading strategies in a lot for shorter forecasting intervals. According to Gehrig and Menkhoff (2006), the realizations of dummy buy or sell signals of different technical trading rules are generated by past price records. Then, traders could use these series of dummy signals to make their trading decision. I collect two unique datasets to make this relevant empirical work. The first key dataset is individual tick-by-tick data. The outstanding point of this data is that it includes traders' identification and record all of their transaction details. Based on this and other elements, I disclose technical traders' different strategies by real market participants. More details of this dataset description are in section 3. Another dataset is market price tick-by-tick data. It is used to generate dummy signals. It includes all price record also as tick-by-tick, high frequency, per second data, which covers all selected underlying futures contracts in the individual transaction dataset. After smoothing this dataset, the dummy trading signals of selected technical trading rules are generated by the smoothing per second price data. In this paper, I select three popular kinds of technical trading rules, which are Momentum, Moving Average, and Trading Range Break-out. Each kind of rules has different parameter settings and generates a technical strategies universe with 13,500 different rules. It is according to intraday trading time in Chinese futures market. Also, I 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, I 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, I employ k-means clustering algorithm to classify technical traders. According to the clustering results, technical traders' strategies can be divided into 11 groups in my dataset. On the basis of coordinates of 11 clusters, the characteristics of technical strategies in each group are disclosed at the end.

The structure of this paper is as below: section 2 describe the details of datasets in this empirical work; section 3 detailed states the process of filtering technical traders and generation and matching of dummy trading signals; section 4 shows the working of k-means cluster algorithm and its results; section 5 introduce a simple regression model to disclose the features of strategies in each group; and section 6 displays a general conclusion of this paper.

2: Data Description:

In the financial market, futures market is also a high-liquidity market as stock and foreign currency. Previous research on this area mostly analyzed on stock index futures, however, most products on futures market are commodities, such as corn and copper. Commodity futures could be more influenced by macroeconomic effect and the risk level is higher than the other financial market due to its attribute of hedging and higher leverage rate (Fabozzi, Fuss, & Kaiser, 2008). Investors should be more careful about their investment than stock market due to the trading assets which may be very important to 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' behavior, the traders would be recognized as pure technical traders. Thus, this paper utilize the comprehensive tick-by-tick transaction data which can fully reflect market participants' trading behavior, 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 broker in China and one famous data statistics company. In 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 transferring participants' order to the main exchange: Shanghai Futures Exchange, Zhengzhou Commodity Exchange, and Dalian Commodity Exchange. The data only covers one main commodity—Rebar (RB), which is trading at Shanghai Futures Exchange.

Even Chinese futures market is very different to other main futures market in the related research, such as LME and CBO, Chinese futures market also is one of the most activity market in the world. Chinese financial derivative has even just started developing in the past 20 years. Until now, financial derivatives have attracted more and more attention and had a speedy promotion. Futures is a relative measure product in Chinese financial derivative market and it plays a quite important role in the global futures market, e.g. Shanghai Futures Exchange is the second biggest exchange for copper trading in the world and Chinese rebar futures is the most traded metal futures currently in the whole world. Thus, I choose rebar contract by some following reasons. Rebar contract launched on 27th March, 2009, and it is still a very active futures contract until now. 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 with lack of rebar. Also, China mainland has a great inventory of iron ore, which is the raw material of rebar. Thus, the demand and supply volumes of rebar are absolutely enough to support a high liquidity trading market. Due to reduce the risk, steel industries and steel trading business, who need or produce rebar, would not only consider the spot market but also invest in the futures market for hedging or arbitrage. This is also an interesting and special futures market that it looks like a pure speculative financial market. Shanghai Futures Exchange doesn't encourage investors delivery real commodities after execution day, which means near all of market participants must close out all of their positions before the end of each contract. Most interesting stories are in other paper about how trader loses their money in the financial market.

The utilized 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 include 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)[1], trading indication (open or offset and long or short)[2], and type-sign of investors (individual or institution). The total records of this database are 4,427,131. Different commodity futures contract may have different number of contracts in one year. For rebar contract, there are 12

[1] Net position means investor will sell or buy how many contracts of commodity on delivery day after one transaction. Real position

[2] In the futures market, long and short means investors take what kinds of position. Actually, taking long position means buy futures contracts. Taking short position means sell futures contracts. Open and offset indicate 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).

contracts in one year. And, each contract starts trading at the beginning of each month and delivery or execute at the same time (15th in each month) in next year. For example, rb201006 started trading on Jun. 16th 2009 and delivered on Jun. 15th 2010. Because rebar contract launched at March, 2009, the first contract is 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 uncompleted records of rb201211 to rb201310. These total 50 contracts establish a different cross-section to different traders. Trader code marks different 19,933 investors. This paper only discusses investors' transaction part. So, the records only content investors' transaction orders and does not cover investor's other bid and ask orders.

The second dataset is the tick-by-tick high frequency data (data2) of the whole market price records. I collect the data from the mentioned data statistic company[3]. 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 transaction 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 have 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 realized by data1.

Meanwhile, this paper also utilizes some main macroeconomics index announcement to identify fundamental and technical traders which the method refers to Jiang, Lo, and Valente (2013). Rebar market is quite sensitive by government macroeconomic policy, because of its main function as I described above. Therefore, the third part of data is about announcement time of macroeconomic index. Whether this public

3 www.gtarsc.com/

information can affect investors' preference and act should be examined. I 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 monthly announced and for EI and GDP is quarterly announced from Chinese government. If traders' trading volume is correlated with some indexes, it means these traders are not pure technical investors. If trader's volumes do not have any relationships to these macroeconomic indices, these groups of traders would be recognized as pure technical traders.

3: Empirical Analysis

3.1: Handling Endogenous Variables and Irrational Transaction

In the first step, I 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.

The endogenous variable is the transaction price, which is unavoidable. In the regression model, I 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 I discussed before, the owned transaction price data should be recognized as an endogenous variable due to the data1 is just a part of the whole market. For instance to indicate, the data is just a part of 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 my owned data. On the one hand, outside of my owned data, investors affect market price change which also means these acts impact on inside sample investors' transaction price. On the other hand, outside sample investors' trading volume can significantly affect inside sample traders' trading volume due to herding (Nofsinger & Sias, 1999). This paper refers the theorem of Instrumental Variable Estimation (IV) to deal with this problem. Based on Wooldridge (2011), if the instrumental variables are absolutely exogenous to the regression model, two-stage least square (2sls) can be dividedly achieved. Thus, I made four kinds of average price rely on the data1&2 as the instrumental variables to transaction price in my research sample: ap1: previously average 1 hour total records' price of broker sample; ap2: previously average 500 records' price of broker sample; ap3: previously average 1 hour 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 4 IVs are 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 97% correlation with transaction price in my original sample. Then, I adopt OLS to get the predict value of transaction price with the four IVs: ap1 to ap4 and the exogenous factor: individual position and other factors. The price-hat (predict value of price) take the place of original price in the regression model. Generally speaking, this method divides 2sls into two steps. It causes different stand error 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, I process the initial data. The initial data of trading volume is nominal one which is how many units of contract trade. Regard rebar futures contract, one trading unit of contract actually is equal to 10 tons of rebar. Thus, I adopt using nominal trading volume to multiply 10 to achieve real trading volume in order to deflate the price effect.

The second problem, I call it "irrational trading behaviour" firstly which 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 delist. The question is from here, if an irrational trader still holds some contracts just before delivery day, he will drop his position even the market price is very unexpected for his portfolio. The reason is that most of this kind of investors do not have real commodities for delivery and also do not have utilization of commodities. It is corresponding with the regulation of Shanghai Futures Exchange. 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, I 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 & 2. I 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, In other words, it means the contract become inactive from active generally before 2 month of the delivery day. Therefore, I move out all the transaction records from data during last two month of each rebar contract. I also exam 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, I utilize 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.

3.2: Multiple Regression Model for Filtering Pure Technical Traders

In this section, I design a reasonable multiple regression model to filter pure technical traders for Chinese rebar futures. The regression 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 the following:

$$\boldsymbol{ln(v) = C + \alpha ln(price) + \beta ln(position) + \sum_{i=1}^{n} d_i(\alpha_i ln(price) + \beta_i ln(position) + \gamma_i \Delta T_{PPI} + \delta_i \Delta T_{PMI} + \varphi_i \Delta T_{REI} + \theta_i \Delta T_{EI} + \rho_i \Delta T_{GDP}) + \varepsilon}$$

Where, $\boldsymbol{position_t = market\ position_t - individual\ net\ position_t}$

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

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

$$\boldsymbol{\Delta t = D_A - D_O}$$

And, $\mathrm{D_{A+1}}$ is the next announcement date of each macroeconomic index, $\mathrm{D_A}$ is the last monthly announcement date, and $\mathrm{D_O}$ is the transaction of each record occurred date.

Where, C is constant item, ε is error term, v is individual trading volume in each transaction, price is the described price-hat instead of transaction price (market price) for each record, position is individual net position which means how many contracts each participant hold implying difference between opened and closed position. In addition, I 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. I 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 dummy setting for different traders which depend on the size of research sample (can 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. The setting of T is the time changing trend between monthly (or quarterly) announcement and next announcement time of each macroeconomic index. 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 five relevant indices cannot influence decided trading volume of pure technical traders. 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, I 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 top most trading 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 trading traders, they should have significance to investigate and summarize the total sample of technical traders. In addition, these top 100 traders are ordered by their amount of 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.

According to the second part, it is the research on single active futures contract. Since, rebar futures contract was starting on March 27th 2009 in Chinese futures market. There are only September, October, November, and December contracts existing in 2009. With comparing trading volume and market position situation of each included contracts. From 2010 to 2012 (also content contracts in 2013), January, May, and October contracts are relatively active contracts. This is caused by the seasonal economic cycle reason in China. Therefore, based on the variation trend of trading volume, market position, and amount of trading record, this part chooses following contracts with boldface letter as the investigated single contract.

Contract Code	Maximum Market Position	Total Trading Records	Contract Code	Maximum Market Position	Total Trading Records
rb0909	**423,426**	**79,028**	**rb1110**	**888,390**	**363,451**
rb0910	397,842	30,407	rb1111	1,534	119
rb0911	**895,870**	**142,224**	rb1112	584	2
rb0912	**1,008,870**	**190,001**	**rb1201**	**772,958**	**142,304**
rb1001	**1,140,896**	**243,295**	rb1202	2,292	2
rb1002	**1,011,436**	**123,746**	rb1203	308	9
rb1003	195,814	19,258	rb1204	556	8
rb1004	53,498	1,745	**rb1205**	**865,430**	**211,069**
rb1005	**1,074,692**	**374,967**	rb1206	1,906	22
rb1006	53,270	1,789	rb1207	0	0
rb1007	9,626	335	rb1208	0	0
rb1008	17,078	1,333	rb1209	2,038	77
rb1009	26,948	1,442	**rb1210**	**959,664**	**161,609**
rb1010	**1,497,516**	**619,286**	rb1211	1,310	13
rb1011	26,292	1,782	rb1212	1,264	54
rb1012	36,544	1,567	**rb1301**	**1,891,202**	**340,220**
rb1101	**1,426,002**	**417,135**	rb1302	3,404	27
rb1102	21,718	427	rb1303	244	14
rb1103	13,024	658	rb1304	0	0
rb1104	13,044	358	**rb1305**	**1,114,852**	**95,830**
rb1105	**1,328,002**	**326,670**	rb1306	0	0
rb1106	1,556	138	rb1307	0	0
rb1107	7,690	114	rb1308	0	0
rb1108	506	4	rb1309	788	14
rb1109	2,746	1,289	rb1310	1,790	38
				total	3,893,880

Thus, these 15 individual contracts, black in the above table, are surveyed in the second part of each single futures contract. The method is same as the first part. But, the investigated traders increase 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, I have also made a secondary task. Based on individual position and immediate transaction price and their product (real transaction value), I 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.

According to the regression results, if all of the five macroeconomic timing variables are not significant to the traders' trading volume, which means the macro or external factors are not able to effect trading behaviour, this group of trader should be recognized as pure technical traders. The initial results show how many technical traders of top 200 most trading traders existing in these 15 samples as the following table:

	Long	Short		Long	Short
rb0909	61	59	rb1105	69	65
rb0911	81	74	rb1110	77	69
rb0912	67	74	rb1201	70	61
rb1001	52	64	rb1205	59	68
rb1002	79	74	rb1210	55	78
rb1005	54	55	rb1301	65	62
rb1010	54	73	rb1305	79	105
rb1101	64	79			

After statistics, there are about 50 traders are recognized 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 trading traders.

3.3: Investigate Sample Selecting

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, I 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. Therefore, I select traders from top

200 most trading traders in each main futures contract. Some of them appear and can be selected in different contract, but I only choose one to symbolize this special traders. For instance, if trader 666 is identified as technical trader in two contracts, I only choose one contract as his research sample. After statistics, I 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' behavior 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. I select three kinds of popular technical trading rules as the bench mark of pure technical traders to investigate their behaviors.

3.4: Selected Technical Trading Rules

This research only selects three popular classes of technical trading rules (Momentum, Moving Average, and Trading Range Breakout). 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: neutral

$n \in \{1, 2, \dots, 13500\}$. (1 day=3.75 hour=225 minutes=13500 seconds)

(Trading time from SHANGHAI Futures Market in one day:

9:00-10:15: Trading (1 hour and 15 minutes)

10:15-10:30: Short Break

10:30-11:30: Trading (1 hour)

11:30-13:30: Noon Break

13:30-15:00: Trading (1 and half an hour)

Total: 3.75 hours

1, Momentum Rule (MO) refers to Conrad & Kaul (1998) 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.

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

Also, based on the momentum rule, I 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).

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

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, Griffioen, and Hommes (2000). 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. I also introduce Contrarian Moving Average Rule (IMA), which have the inverse signals to MA.

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

$$MA_t = \sum_{t=t-n}^{t} P_t/n + 1$$

3, Trading Range Breakout Rule (BO) is generally known as price channel systems. I refer to the literature from Park and Irwin (2005). 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 (2013), the principle of BO is to utilize the local maximum and minimum price as the motivation of technical traders in order to implement their strategies. Based on different period, trader seeks the extreme price as their "support", and they trust the price trend will follow this "support". I still introduce the Contrarian Trading Range Breakout Rule (IBO) as the same principle of CMO and CMA.

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

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

Before the first breakout, the indicator always keeps 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.

3.5: 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 modify (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 utilizes 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, I use average 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, I utilize data2, which includes all ticks for all contracts, to smooth in order to generate dummy trading signals under different technical rules. I refer the generally average method from Simonoff (1998), which is "unweighted sliding-average smoothing method. This algorithm is described as below: I 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 "$P_M \; and \; P_N$". "N-M=T" is the time difference and "$P_M - P_N = P_T$" is the price change between the two time points. Thus, the price movement per unite time between "N and M" is: "$\Delta P = P_T / T$". Through continually adding "ΔP", all time point will have a price record, and it follows average smoothing principle:

$$P_M = P_N - T * \Delta P$$

$$P_{M+1} = P_M + \Delta P$$

$$P_{M+2} = P_M + 2 * \Delta P$$

$$\vdots$$

$$P_{N-2} = P_N - 2 * \Delta P$$

$$P_{N-1} = P_N - \Delta P$$

$$P_N = P_M + T * \Delta P$$

For example, the following table is the effect picture of data2 before and after smoothing algorithm:

time	price
30/10/2009 13:30:02	4227
30/10/2009 13:31:39	4236
30/10/2009 13:36:49	4247
30/10/2009 13:36:54	4247
30/10/2009 13:41:35	4240
30/10/2009 13:42:15	4238
30/10/2009 13:45:58	4245
30/10/2009 13:47:40	4244
30/10/2009 13:48:02	4245
30/10/2009 13:50:14	4245
30/10/2009 13:50:27	4249
30/10/2009 13:52:29	4237
30/10/2009 13:52:58	4236
30/10/2009 13:52:59	4236
30/10/2009 13:56:18	4236

Time	Price
30/10/2009 13:30:02	4227
30/10/2009 13:30:03	4227.09
30/10/2009 13:30:04	4227.19
30/10/2009 13:30:05	4227.28
30/10/2009 13:30:06	4227.37
30/10/2009 13:30:07	4227.46
30/10/2009 13:30:08	4227.56
30/10/2009 13:30:09	4227.65
30/10/2009 13:30:10	4227.74
30/10/2009 13:30:11	4227.84
30/10/2009 13:30:12	4227.93
30/10/2009 13:30:13	4228.02
30/10/2009 13:30:14	4228.11
30/10/2009 13:30:15	4228.21
30/10/2009 13:30:16	4228.3
30/10/2009 13:30:17	4228.39
30/10/2009 13:30:18	4228.48
30/10/2009 13:30:19	4228.58
30/10/2009 13:30:20	4228.67
30/10/2009 13:30:21	4228.76
30/10/2009 13:30:22	4228.86
30/10/2009 13:30:23	4228.95
30/10/2009 13:30:24	4229.04
30/10/2009 13:30:25	4229.13
30/10/2009 13:30:26	4229.23

The time series (time division) data has been created in the last step so that I start completing the generation of dummy trading signals with different rules. Due to the huge data size, I utilize C++ to create all dummy signals. All algorithms and computational programs are realized by Microsoft Visual C++6.0 and QT 5.1.1. For the rules, I select 6 classes of rule, and each class of rules includes 13,500 types by different parameters. So, the amount of rules is 6*13500=81,000 in the universe. Also, the main research contracts are 15 mentioned active contracts previously, thus I split 15 contracts as individual contract to generate signals with 81,000 rules. Then, I 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.

Now, I have produced and introduced my data3, which is very important to explain the effect of all selected technical rules. The next work is how to link data3 to data1 in order to investigate pure technical traders' strategies. It is difficult to investigate every trader's specific behavior so that I trend to classify different types of technical traders. In each type, members should have generally similar strategies. In this section, I describe my method to link data1 and data3 and also show adopted classification method, which is K-means clustering algorithm.

4: Trader Classification

4.1: Data Reconstitution

There is a connection between data1 and data3. I adopt a simple and sensible method, which is to calculate the similarity (%) between 81 traders' real actions and dummy signals of each rule. In data3 as I 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 amount of transactions (observations or actions).

I firstly filter 81 traders' data from data1 to create 81 individual datasets. Next, I 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 added in the reconstituted dataset. Each column indicates the dummy trading signals of one specific technical rule. Then, I 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, I can utilize 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, I achieve a significant matrix, and the size is 81,001*82 (include title row and first column, in fact, it is 81,000*81), which covers 81 traders' similarity to all 81,000 rules. I put a short sample as below:

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%	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	…

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 "female parent" of traders' classification, because the similarities traders' different preference of each selected technical rules. Thus, I classify 81 traders in different groups based on the above mentioned similarity.

4.2: K-Means Clustering Algorithm

Cluster analysis[4] 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, I trend 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 attribute in the clustering space so that clustering algorithm is easy and sensible to realize 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 traders' different preference of technical rules. Therefore, the above mentioned space is an 81,000-dimentional 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 I 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, I utilize Matlab R2012b to realize k-means clustering because Matlab has

[4] For the Information and knowledge of clustering analysis, I refer to "Maimon, O. Z., & Rokach, L. (2010), Data Mining and Knowledge Discovery Handbook. New York, Springer".

standard procedure package of k-means. Also, I adopt the default distance—squared Euclidean distance (SED) of this automatic procedure. More details about the principle of SED are in appendix 1. 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 3 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 I can get a sensible classification of traders.

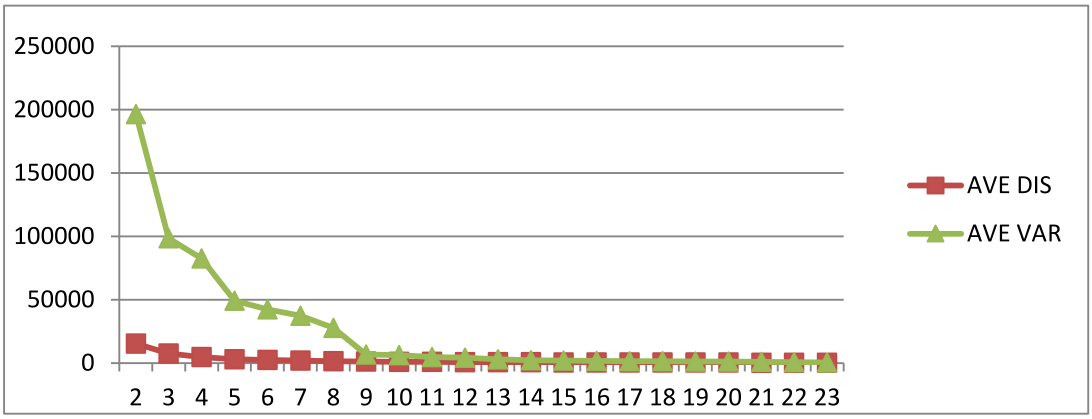
There is one significant problem is that how to decide the number of k. We cannot randomly set a number of k with our "intuition". There are many methods discussing in clustering area. In this paper, I refer to the method of "low average sum of variance and distance" (Maimon & Rokach, 2010). The principle 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, I designed three projects (three samples) to proof the correct number of k.

4.3: Clustering Results

In project one, I put the 81,000*81 matrix in the k-means algorithm and get the following statistic results, figure 4.1. Where, the x-axis is the number of k. I make 23 times of clustering with setting k equal 2 to 23. The y-axis is the value of average sum of distance and variance[5]. 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

[5] 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, I 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.

become stable. Thus in project one, all 81 traders should be divided into 11 groups with 81,000 attributes (rules).



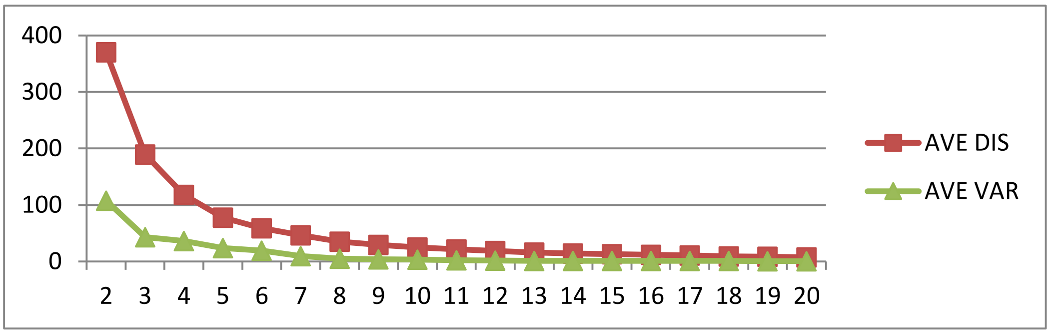
Project 1: It initially proofs the number of classifications in the research, and it covers all investigated rules. However, this is a biased estimation, and the dimensional-space is very complex. Thus, I designed other two projects to support the clustering results. The principle is that I reduce the dimension of attributes in the space—I 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 is in appendix 2 (286*6=1716 rules). The feature is that choosing a rule every 5, 10, 15, 30 seconds before 5 minutes end, and then choosing a rule every 50 seconds until one day end (13,500 seconds).

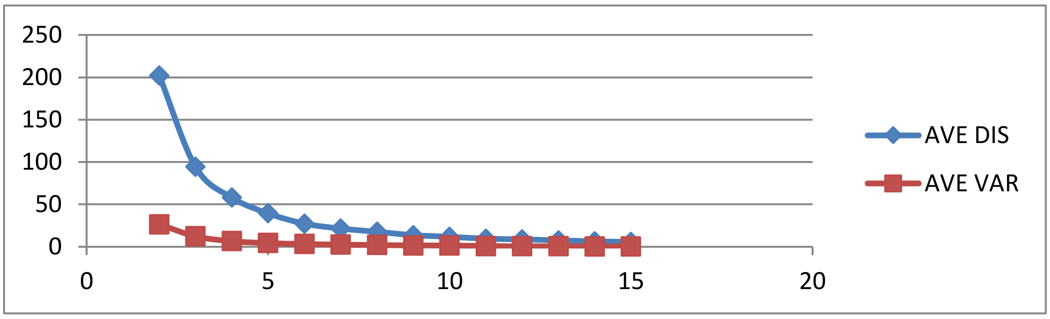
Project 3: Similar to P2, but it refers to the previous research. It has fewer rules to cluster, and the interval setting is in appendix 3 (90*6=540 rules). The feature is that choosing a rule every 5, 10, 15, 30, 60 (1 min.), 600 (10 min.), 1200 (20 min.) seconds with increasing time.

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 recognized as the correct number of k. The below two graphs show the clustering results of project two and three. X-axis and y-axis have the same explanations of project one.

P2:



P3:



I make 20 operations of project two for the number of k equals 2 to 20 and 15 operations of project three for the number of clusters equals 2 to 15. Also, due to the k-means algorithm randomly set the centroids, I 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 in the following table, I use 1 to 81 to label 81 traders:

Group	P1:	P2:	P3:
1	1, 8, 9, 16, 20, 22, 24, 33, 34, 37, 38, 44, 46, 48, 50, 53, 62, 63, 74	1, 8, 16, 20, 22, 24, 31, 33, 34, 37, 38, 44, 46, 48, 50, 53, 60, 62, 63, 76	54, 66, 72
2	15, 26, 57, 68, 71, 73	15, 26, 57, 68, 71, 73	39, 41, 42, 51, 77
3	11, 21, 40, 41, 58, 75, 77, 79	54, 66, 67, 69, 72	2, 5, 15, 26, 55, 57, 65, 70, 71, 73
4	3, 4, 12, 23, 27, 28, 36, 43, 51, 67, 69, 80	10, 21, 40, 58, 75, 77, 79	18
5	25, 39, 54, 66, 72	4, 11, 12, 23, 25, 27, 36, 39, 41, 43, 51, 80	4, 11, 75, 79
6	18	18	33, 47, 58, 62, 80
7	10, 59, 61	59, 61, 74	1, 3, 8, 20, 24, 27, 28, 34, 38, 50, 53, 61, 63, 67, 69
8	42, 47, 78	42, 47, 78	6, 9, 13, 14, 16, 17, 19, 29, 30, 32, 52, 64, 68, 74, 81
9	13, 14, 17, 19, 29, 30, 32, 52, 64, 81	3, 6, 9, 13, 14, 17, 19, 28, 29, 30, 32, 52, 64, 81	22, 31, 37, 44, 48, 59, 60, 76
10	2, 5, 6, 31, 55, 60, 65, 70, 76	2, 5, 55, 65, 70	10, 12, 21, 23, 36, 40, 43, 46, 78
11	7, 35, 45, 49, 56	7, 35, 45, 49, 56	7, 25, 35, 45, 49, 56

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).

5: Empirical Results

The key point of clustering is to use the centered coordinates to describe members' attribute in each group. After clustering process of project one, 81 pure technical traders are divided into 11 groups. Thus, the 11 centroids summarize characteristics of members in their individual set. I operate the algorithm with setting k=11 about 50 times to achieve relatively lest ASD and ASV. The results of coordinates construct an 11*81,000 matrix, where 11 rows imply 11 centroids and 81,000 columns imply similarity degree with 81,000 technical rules. Using this matrix, I 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, I select higher similarity rules and return to the dataset of dummy signals (4.2.2) to draw out dummy signals according to the specific rules for every trader. Then, I make a simple regression between each trader's real action and dummy signals with rules of higher similarity, which is in order to support the relationship between specific rules and traders' strategies.

The selected technical rules cover six kinds of rules as mentioned before. Also, I extend the time length to 13,500 seconds, which is trading time in one day, for each variety. Thus, the amount rules are 13,500*6=81,000 as above discretion. I adopt a simple regression model to investigate the effect of rules to traders' real actions as following:

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

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 13500 rules. Constant term is c and error term is ε. The independent variables are all dummy signals with all 81000 rules. Thus, I 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 realized due to the huge similar signals with different rules. In other words, the problem of col-linearity happens. Each signals' variable only include 1, 0, and -1. Same or correlated variables possibly exist in total 81000 rules (after experiment, same rules exist). The general methods of this question are to remove the same or correlated variables in the model. But, it is not advisable in this research. For example, if the vector of BO22 is equal to MA16 for trader 1, I don't know which rule we need to remove. If I remove both two rules, it influences the regression results because MA16 (or BO22) may be very significant to trader 1's real trading. So, I stop investigating all 81000 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, choosing 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:

$$S_{jg} = c + MA_{rg} + MO_{rg} + BO_{rg} + IMA_{rg} + IMO_{rg} + IBO_{rg} + \epsilon$$

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 13500. As I described before, based on the 11 clusters' coordinate, six rules of MA, MO, BO, IMA, IMO, and IBO with the greatest similarity are able to select in each group. They are components of T6's multiple regression. I 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, I 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. I produce the following tables to summarize the final results.

Group	Amount of Traders	Highest Six Rules	Significant Traders	Significant Probability	Insignificant Traders	Insignificant Probability
1	19	bo13	17	89.47%	2	10.53%
		ma312	13	68.42%	6	31.58%
		mo251	12	63.16%	7	36.84%
		ibo990	10	52.63%	9	47.37%
		ima6355	9	47.37%	10	52.63%
		imo13406	8	42.11%	11	57.89%
2	6	bo351	5	83.33%	1	16.67%
		ma539	6	100.00%	0	0.00%
		mo387	6	100.00%	0	0.00%
		ibo3573	6	100.00%	0	0.00%
		ima9094	4	66.67%	2	33.33%
		imo6375	4	66.67%	2	33.33%
3	8	bo7320	3	37.50%	5	62.50%
		ma11723	4	50.00%	4	50.00%
		mo11973	5	62.50%	3	37.50%
		ibo122	7	87.50%	1	12.50%
		ima296	4	50.00%	4	50.00%
		imo235	6	75.00%	2	25.00%
4	12	bo1	10	83.33%	2	16.67%
		ma18	10	83.33%	2	16.67%
		mo13136	9	75.00%	3	25.00%
		ibo113	9	75.00%	3	25.00%
		ima11694	7	58.33%	5	41.67%
		imo7583	9	75.00%	3	25.00%
5	5	bo8	5	100.00%	0	0.00%
		ma41	5	100.00%	0	0.00%
		mo13114	5	100.00%	0	0.00%
		ibo453	5	100.00%	0	0.00%
		ima1136	3	60.00%	2	40.00%
		imo556	4	80.00%	1	20.00%
6	1	bo506	1	100.00%	0	0.00%
		ma3503	1	100.00%	0	0.00%
		mo1297	1	100.00%	0	0.00%
		ibo13476	0	0.00%	1	100.00%
		ima10	1	100.00%	0	0.00%
		imo9780	0	0.00%	1	100.00%
7	3	bo13500	1	33.33%	2	66.67%
		ma13383	1	33.33%	2	66.67%
		mo12971	2	66.67%	1	33.33%
		ibo369	2	66.67%	1	33.33%
		ima1011	3	100.00%	0	0.00%
		imo887	2	66.67%	1	33.33%
8	3	bo2484	1	33.33%	2	66.67%
		ma5232	2	66.67%	1	33.33%
		mo3071	2	66.67%	1	33.33%
		ibo94	3	100.00%	0	0.00%
		ima212	2	66.67%	1	33.33%
		imo275	2	66.67%	1	33.33%
9	10	bo32	6	60.00%	4	40.00%
		ma65	10	100.00%	0	0.00%
		mo92	8	80.00%	2	20.00%
		ibo5425	9	90.00%	1	10.00%
		ima13393	7	70.00%	3	30.00%
		imo11704	6	60.00%	4	40.00%
10	9	bo40	8	88.89%	1	11.11%
		ma215	8	88.89%	1	11.11%
		mo193	8	88.89%	1	11.11%
		ibo11156	6	66.67%	3	33.33%
		ima12324	4	44.44%	5	55.56%
		imo13146	7	77.78%	2	22.22%
11	5	bo13227	5	100.00%	0	0.00%
		ma20	4	80.00%	0	0.00%
		mo13414	2	40.00%	3	60.00%
		ibo68113	4	80.00%	1	20.00%
		ima2170	5	100.00%	0	0.00%
		imo536	5	100.00%	0	0.00%

Table explanation: 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 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 utilize 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 utilize 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 utilize 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.

6: Conclusion:

This is an empirical research on technical trading strategies. The contribution of this research is to indicate technical trading behavior for real traders in Chinese futures market. I select rebar futures contracts, which is one of main 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. Top 100 most trading traders are the main research object due to they are more likely to be the technical traders and employ program trading. I choose five related macroeconomic indexes to rebar market as the filter factor by using a simple multiple-regression model to filter technical traders. The results shows, in each contract, technical traders occupy about 20% from top 100 most trading traders. Because it is high frequency per second data, I use similar tick-by-tick data, which records complete market price trend of each contract without traders' identification, to generate dummy signals with a series of technical trading rules. Then, dummy signals of each selected technical rules and each trader's real action are combined and matched according to same transaction time. I calculate the similarity between them which indicate each trader's motivation to employ each rule. I only select 81 technical traders from 15 most active contracts in my dataset to investigate. Based on the similarity matrix, I adopt k-means clustering algorithm to classify these 81 traders. The clustering results show that they can be divided into 11 groups with different technical strategies. In order to avoid same or correlated dummy signals with different rules, I choose top 6 highest rules based on the coordinates of 11 clusters to state whether these 6 rules are significant to trader's real actions. The results indicate that most members, in the different groups, must have one or more significant technical rules to their real action. More details are displayed in section 5.

Reference:

Bagehot, W. and Treynor, J., 1971, the Only Game in Town, Financial Analysts Journal 27, 2, 12-14

Bodie, Z., 1983, Commodity Futures as a Hedge against Inflation, Journal of Portfolio Management Spring 12–17

Bodie, Z., Rosansky, V., 1980, Risk and Returns in Commodity Futures, Financial Analysts Journal May/June, 27–39

Boswijk, P., Griffioen, G., & Hommes, C., 2001, Success and Failure of Technical Trading Strategies in the Cocoa Futures Market, (Tinbergen Institute Discussion Paper 01-016/1), Tinbergen Institute Amsterdam and Rotterdam

Brock, W., Lakonishok, J., LeBaron, B., 1992, Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, Journal of Finance 485, 1731–1764

Chan, K., Hameed, A., Tong, W., 2000, Profitability of Momentum Strategies in the International Equity Markets, Journal of Financial and Quantitative Analysis 35,153–172

Conrad, J., Kaul, G., 1998, an Anatomy of Trading Strategies, Review of Financial Studies 11, 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, L. J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J., 1990, Noise Trader Risk in Financial Markets, The Journal of Political Economy, 98, 4, 703-738

De, L. 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, High Finance in Copper, Financial Analysts Journal (Nov/Dec), 133–142

Erb, C., Harvey, C., 2006, the Strategic and Tactical Value of Commodity Futures, Financial Analysts Journal 62 (2), 69–97

Faber, M., 2007, A Quantitative Approach to Tactical asset allocation, Journal of Wealth Management 9, 69–79

Fabozzi, F. J., Füss, R., & Kaiser, D. G. (2008), *The Handbook of Commodity Investing*, Hoboken, N.J: John Wiley & Sons

Fama, E. F., 1970, Efficient Capital Markets: A Review of Theory and Empirical Work. (The journal of finance; vol. 25, no. 2, 1970) Illinois, Chicago: Univ., Center for Mathematical Studies in Business and Economics

Fifield, S.G.M, D.M. Power and D.G.S Knipe, 2008, “The performance of moving averagerules in emerging stock markets” Applied Financial Economics, Vol. 18, No. 19, October 2008, pp. 1515

Foucault, T., Kadan, O., Kandel, E., & Centre for Economic Policy Research (Great Britain), 2001, Limit order book as a market for liquidity. London: Centre for Economic Policy Research

Gehrig, T., Menkhoff, L., 2006, Extended Evidence on the Use of Technical Analysis in Foreign Exchange, International Journal of Finance and Economics 11, 327–338

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

Gençay, R., 1999, Linear, non-Linear and Essential Foreign Exchange Rate Prediction with Simple Technical Trading Rules, Journal of International Economics, 47, 1, 91-107

Gorton, G., Rouwenhorst, K., 2006, Facts and Fantasies about Commodity Futures, Financial Analysts Journal 622, 47–68

Hull, J., 2012, Options, Futures, and other Derivatives, Essex, England: 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., (2013), Market Ecologies: The Interaction and Profitability of Technical Trading Strategies, Discussion Papers in Economics 13/02, Department of Economics, University of Leicester

Jegadeesh, N., Titman, S., 1993, Returns to Buying Winners and Selling Losers: Implications for stock market efficiency. Journal of Finance 48, 65–91

Jiang, G.J., Lo, I., Valente, G., 2012, High Frequency Trading in the US Treasury Market: Evidence Around Macroeconomic News Announcements. Working Paper, Washington State University

Lehmann, B., 1990, Fads, Martingales, and Market Efficiency, Quarterly Journal of Economics 105, 1–28

Levich, R.M. and Lee R. T., 1993, the Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach, Journal of International Money and Finance, Vol. 12, pp451-474

Lo, A., Mamaysky, H., Wang, J., 2000, Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, Journal of Finance 55, 1705–1770

Lo, A.W., MacKinlay, A.C., 1990, When are Contrarian Profits due to Stock Market Overreaction? Review of Financial Studies 3, 175–208

Lucke, B., 2003, Are Technical Trading Rules Profitable? Evidence for Head-and-Shoulder Rules, Applied Economics, Vol. 25, No.1, pp33-40

Lui, Y., Mole, D., 1998, the Use of Fundamental and Technical analysis by Foreign Exchange Dealers: Hong Kong Evidence. Journal of International Money and Finance 17, 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, 623–640

Maimon, O., & Rokach, L., 2010, Data Mining and Knowledge Discovery Handbook, Dordrecht: Springer

Marshall, B.R., Cahan, R.H., Cahan, J.M., 2008, Does Intraday Technical Analysis in the US Equity Market has Value? Journal of Empirical Finance

Miffre, J., Rallis, G., 2007, Momentum Strategies in Commodity Futures Markets, Journal of Banking and Finance 31 (6), 1863–1886

Neely, C., Weller, P., Ulrich, J., 2009, the Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market, Journal of Financial and Quantitative Analysis 44, 467–488

Neely, Christopher J. and P. A. Weller, 2003, Intraday Technical Trading in the Foreign Exchange Market, Journal of International Money and Finance, Vol. 22, No.2, pp223-237

Neely, Christopher J., 2002, The Temporal Pattern of Trading Rule Returns and Exchange Rate Intervention: Intervention Does Not Generate Technical Trading Profits, Journal of International Economics, Vol. 58, No. 1, pp211-232

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, 81–93

Park, C.-H., Irwin, S.H., 2004, the Profitability of Technical Analysis: A review. AgMAS 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

Simonoff, J. S., 1998, Smoothing Methods in Statistics. New York: Springer-Verlag

Slezak, S. L., 2003, On the Impossibility of Weak-Form Efficient Markets, The 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 245, 1647–1691

Sweeney, Richard J., 1986, Beating the Foreign Exchange Market, Journal of Finance, Vol. 41, pp163-182

Tian, G., Wan, G. & Guo, M., 2002, Market Efficiency and the Returns to Simple Technical Trading Rules: New Evidence from U.S. Equity Market and Chinese Equity Markets. Research Online

Vrugt, E.B., Bauer, R., Molenaar, R., Steenkamp, T., 2004, Dynamic Commodity Timing Strategies, Working paper, Maastricht University

Wang, C., Yu, M., 2004, Trading Activity and Price Reversals in Futures Markets, Journal of Banking and Finance 28, 1337–1361

Wooldridge, J. M., 2011, Econometric Analysis of Cross Section and Panel Data, Cambridge, Mass: MIT

Appendix 1: Principle of K-Means Clustering with Squared Euclidean Distance.

I 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 labeled 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 2 original centroids in the space, and their coordinates are "$ce_{1,n,x}, ce_{2,n,x}$". 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:

$$d\left(ce_{1,n,x}, p_{c,1,x}\right) = \sum_{i=1}^{x} (ce_{1,n,i} - p_{c,1,i})^2$$

$$d\left(ce_{1,n,x}, p_{c,2,x}\right) = \sum_{i=1}^{x} (ce_{1,n,i} - p_{c,2,i})^2$$

For centroid two to point 3 and 4:

$$d\left(ce_{2,n,x}, p_{c,3,x}\right) = \sum_{i=1}^{x} (ce_{2,n,i} - p_{c,3,i})^2$$

$$d\left(ce_{2,n,x}, p_{c,4,x}\right) = \sum_{i=1}^{x} (ce_{2,n,i} - p_{c,4,i})^2$$

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

$$Average\ Distance\ of\ Group\ 1\ (avdg1) = \frac{d\left(ce_{1,n,x}, p_{c,1,x}\right) + d\left(ce_{1,n,x}, p_{c,2,x}\right)}{2}$$

$$Average\ Distance\ of\ Group\ 2\ (avdg2) = \frac{d\left(ce_{2,n,x}, p_{c,3,x}\right) + d\left(ce_{2,n,x}, p_{c,4,x}\right)}{2}$$

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:

$$Optimal\ Average\ Distance\ of\ Group1 = min.\,(avdg1)$$

$$Optimal\ Average\ Distance\ of\ Group2 = min.\,(avdg2)$$

The finally optimal centroids are captured, and their coordinates is identified with x dimensions. These coordinates is very significant because it is recognized 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.

Appendix 2: Interval Setting of Project 2

	sec.	min.	hour.	day (3.75h per day)
start 5 sec. 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 sec. interval	30	0.500	0.008	0.002
	40	0.667	0.011	0.003
	50	0.833	0.014	0.004
start 15 sec. 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 sec. 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
start 50 sec. 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
	.	.	.	.

Appendix 3: Interval Setting of Project 3

	sec.	min.	hour.	day (3.75h per day)
start 5 sec. 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 sec. 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
	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 sec. 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
start 30 sec. 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
	870	14.500	0.242	0.064
start 1 min. interval	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
	1680	28.000	0.467	0.124
	1740	29.000	0.483	0.129
start 10 min. interval	1800	30.000	0.500	0.133
	2400	40.000	0.667	0.178
	3000	50.000	0.833	0.222
	3600	60.000	1.000	0.267
start 30 min. interval	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
	11700	195.000	3.250	0.867
one day	13500	225.000	3.750	1.000