Order Imbalance Based Strategy in High Frequency Trading

Darryl Shen

Linacre College

University of Oxford

A thesis submitted in partial fulfillment of the MSc in

Mathematical and Computational Finance

May 27, 2015

Acknowledgements

I would like to thank my supervisor, Dr. Zhaodong Wang, for his assistance on this thesis and providing guidance throughout. Discussions about this project, high frequency trading, hedge funds, and the industry as a whole have been interesting and insightful.

From the MSc Mathematical and Computational Finance class, I would like to thank Ivan Lam and Xuan Liu for their ideas and perspectives on high frequency trading, trading strategies, and statistical analysis.

I would also like to thank a good friend in the industry, and all around genius, Jethro Ma.

Finally, my sincerest gratitude to my fiancée Emily and my parents for their unending support.

Abstract

This thesis aims to investigate the performance of an order imbalance based trading strategy in a high frequency setting. We first analyze the statistical properties of order imbalance and investigate its capabilities as a trading strategy motivated by ideas introduced in [4, 7, 11]. We try to understand how the strategy performs on different futures contracts and its relationship with trading volume. Finally, we attempt to improve the trading strategy by including other imbalance-based signals, adjusting for bid-ask spread, and optimizing the model and trading parameters.

Contents

Chapter 1

Introduction

1.1 High Frequency Trading

Traditionally, financial markets operated on a quote-driven process where a few market makers provided the sole liquidity and prices for financial assets [6]. Recently, major developments have been made to electronify the financial markets which has led to many trading firms using computer algorithms to trade financial assets as reported by Wang [14] and Aldridge [1]. *High frequency trading* (HFT), in particular, has been a major topic due to the features that distinguishes it from electronic and manual trading. This includes the extremely high speed of execution (microseconds), multiple executions per session, and very short holding periods (usually less than a day).

Many algorithmic trading strategies have been developed on the advent of high frequency trading coming to the markets. According to Wang [14] and Aldridge [1], the advantages to having computers execute strategies include: higher accuracy, no emotion, lower costs, and technological innovation as the speed of trading becomes greater. Furthermore, by using the available market data, high frequency traders are able to come up with strategies which identify and trade away temporary market inefficiencies and price discrepancies. In this paper, we will be adapting and testing an existing strategy for HFT and verifying its stability and profitability.

1.2 Limit Order Books and Microstructure

Limit Order Books (LOB) allow any trader to become a market maker in the financial markets (Gould *et al.* [6]). It is a mechanism which allows traders to submit limit buy (sell) orders for the asset and the prices they wish to pay (receive). The limit order book is a complex system and understanding it can give insight into traders' intentions and a way to develop trading strategies using the rich and granular data it stores. We will define a few technical terms relating to LOBs that will be used throughout the paper including fields specific to the dataset we will be examining.

The LOB is essentially a matching engine for buyers and sellers in the market [6]. Within a LOB, the best *bid* (*ask*) *price* is the highest (lowest) price a market maker is willing to buy (sell) the asset at to market takers. The maximum number of contracts that the market makers are willing to buy (sell) at the bid (ask) price is called the best bid (ask) *volume*. Any market taker who wishes to buy (sell) at the *counterparty price* can submit a *market order* to trade at the best ask (bid) price up to the ask (bid) volume available. If the market order to buy (sell) is larger than the ask (sell) volume, then they will walk the book; the market taker will continue buying (selling) at the next-best ask (bid) price until their entire market order is filled.

In this project, the data we will use is the China Financial Futures Exchange (CFFEX) CSI 300 Index Futures (IF). It comprises of snapshots taken every 500 milliseconds. From this point on, every time step is in intervals of 500 ms. That is, time $t + 1$ is 500 ms after time t. The tick size of the IF contracts is 0.2 and the tick value is 300 Chinese Yuan (CNY). The trading hours of the contracts on CFFEX is from 9:15 to 11:30 for the morning session, and 13:00 to 15:15 for the afternoon session. A sample of the data for January 16th, 2014 is shown in Table 1.1 below.

Instrument ID	Update time	Volume	Turnover	Open interest	Bid price	Bid volume	Ask price	Ask volume	Second of day
IF1401	9:27:06.0	14589	9.69e9	60011	2213.4	23	2213.8	70	34026
IF1402	9:27:06.0	6337	4.22e9	28960	2218.8	4	2219	53	34026
IF1401	9:27:06.5	14593	9.69e9	60010	2213.4	21	2213.8	70	34026
IF1402	9:27:06.5	6351	4.23e9	28974	2218.8	4	2219	39	34026
IF1401	9:27:07.0	14595	9.70e9	60010	2213.4	22	2213.8	70	34027
IF1402	9:27:07.0	6351	4.23e9	28974	2218.8	6	2219	39	34027

Note: certain fields omitted to save space

Table 1.1: Sample data set for IF1401 and IF1402 on Jan 16th, 2014.

The data provided is in comma separated values (CSV) format and each file presents a single trading day. However, on the CFFEX, two different futures contracts are traded: the CSI 300 Stock Index Futures (IF) and the Treasury Bond Futures (TF). We will only be focusing on the IF contracts for this paper. IF contract maturity is on the third Friday of every month.

- **Instrument ID**: the unique identifier of the futures contract being traded. It begins with IF or TF and followed by a 4-digit integer. The first two digits represents the year and the last two represents the month of contract maturity. For example IF1401 is the IF contract maturing in January 2014.

- **Update time**: the exact time the LOB snapshot was taken, up to 500 millisecond precision.
- **Volume**: the transaction volume of contracts traded since market open (9:15)
- **Turnover**: the CNY-denominated volume traded since market open (9:15). This quantity is calculated by number of contracts $\times$ price $\times$ tick value.
- **Open Interest**: the number of contracts traded that create an open position (not trade to close)
- **Bid/Ask price**: the highest/lowest price a market maker is willing to buy/sell the futures contract at. Equivalently, it is the best price that a market taker can sell/buy the contract at.
- **Bid/Ask volume**: the number of contracts available at the current best bid/ask price.
- **Second of day**: number of seconds (rounded down) since midnight of the trading day.

Another measure we will be using throughout the paper is the mid-price, denoted M_t, which is the arithmetic average of the bid and ask prices at time t. More details regarding the structure and intricacies of the CFFEX can be found in Wang [14].

1.3 Stationarity

As mentioned in Wang [14], high frequency trading and applications of their strategies are closely related to the ergodic theory of stationary processes. We first explain the two types of stationarity for a time series X_t as defined by Tsay [12] and Wang [14]. A time-series X_t is *strongly stationary* if $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ has the same distribution as $(X_{t_1+a}, X_{t_2+a}, ..., X_{t_n+a})$ for all a and any arbitrary integer $n > 0$. A less strict definition of stationarity for the time-series X_t is called *weakly stationary* if the first two moments do not change over time. That is, $\mathbb{E}[X_t] = \mathbb{E}[X_{t+a}] = \mu$ and $Cov(X_t, X_{t+a}) = \gamma_a$ for all time t and arbitrary a. The covariance for any interval a should only depend on a for the process to be weakly stationary.

Strong stationarity is difficult to verify empirically and therefore any hypothesis tests we conduct in this paper will be to verify weak stationarity, including both the Augmented Dickey-Fuller test and the KPSS test. It is sufficient for the data to be weak stationary for traders to build an algorithmic trading strategy which generates positive expected profits to be applied repeatedly and steadily accumulate positive returns [14]. Further theory regarding stationary processes, the strong ergodic theorem, and its applicability in high frequency trading can be found in [10, 14].

1.4 Order Imbalance

Many studies have been conducted to describe the relationship between trade activity (volume) and price change and volatility (see Karpoff [8] for example). As traders submit limit orders to buy (sell), they impact the bid (ask) volumes of the limit order book and thereby gives us a view of the traders' intentions. Categorizing trade volume as either taking the bid (ask) would allow us to gain insight into the direction of the upcoming price changes. To quantify this intent to trade, we look at the difference between the bid and ask volume, called order imbalance. Chordia and Subrahmanyam [4] have found the positive relationship between order imbalance and daily returns on a sample of stocks from the New York Stock Exchange.

Order imbalance is an important descriptor that allows us to understand the general sentiment and direction the market is headed. If informed traders have information that has not been incorporated into the asset price yet, they can take a long (or short) position given the positive (or negative) news and subsequently increasing the imbalance on the asset [7]. Other market participants, who are merely observing this phenomenon in the LOB, would be able to use this information and develop a strategy to generate positive returns.

The next chapter will carefully analyze the relationship between order imbalance and mid-price changes and determine whether it can be used to predict future price changes at the high frequency level. We will also investigate the statistical properties of order imbalance and how they can be applied to a trading strategy to generate statistically significant positive returns on a daily basis. Given that the previous studies by Chordia [3, 4] and Huang [7] were done on much longer time-scales (daily and 5-15 minute intervals respectively), we will verify if the order imbalance theory they presented are still applicable to high frequency data.

Chapter 2

Order Imbalance Strategy

We will be implementing and testing a similar imbalance-based strategy as outlined by Chordia and Subrahmanyam [4], Huang *et al.* [7], and Ravi *et al.* [11] where we enter into a long position when order imbalance is positive and a short position when order imbalance is negative.

2.1 Volume Order Imbalance

The order imbalance in [4, 7, 11] is defined using Lee and Ready's algorithm [9] to classify trades as either buyer-initiated or seller initiated. This is done by checking if the trade price was closer to the bid (sell) or ask (buy) of the quoted price. Rather, our definition is more similar to the *Order Flow Imbalance* used by Cont *et al.* [5] which we will call *Volume Order Imbalance* (VOI):

$$OI_t = \delta V_t^B - \delta V_t^A \tag{2.1}$$

where

$$\delta V_t^B = \begin{cases} 0, & P_t^B < P_{t-1}^B \\ V_t^B - V_{t-1}^B, & P_t^B = P_{t-1}^B \\ V_t^B, & P_t^B > P_{t-1}^B \end{cases}, \qquad \delta V_t^A = \begin{cases} V_t^A, & P_t^A < P_{t-1}^A \\ V_t^A - V_{t-1}^A, & P_t^A = P_{t-1}^A \\ 0, & P_t^A > P_{t-1}^A \end{cases} \tag{2.2}$$

where V_t^B and V_t^A are the bid and ask volumes at time t respectively and P_t^B and P_t^A are the best bid and ask prices at time t respectively. If the current bid price is lower than the previous bid price, that implies that either the trader cancelled his buy limit order or an order was filled at P_{t-1}^B. As we do not have a more granular order or message book, we cannot be certain of the trader intent, hence we conservatively set $\delta V_t^B = 0$. If the current bid price is the same as the previous price, we take the difference of the bid volume to represent incremental buying pressure from the last period. Lastly, if the current bid price is greater than the previous price, we can interpret this as upward price momentum due to the trader's intent to buy at a higher price. Downward price momentum and sell pressure can be interpreted analogously from the current and previous ask prices.

Order imbalance has positive autocorrelation as presented in Figure 2.1. For most days, the order imbalance autocorrelation is significant up to lag 15. Its first difference has a significant lag-1 negative autocorrelation and is consistent with the results by Chordia [3]. This indicates that positive (negative) imbalances are often followed by periods of persistent positive (negative) imbalances due to traders splitting their orders across multiple periods as explained by Chordia [4].

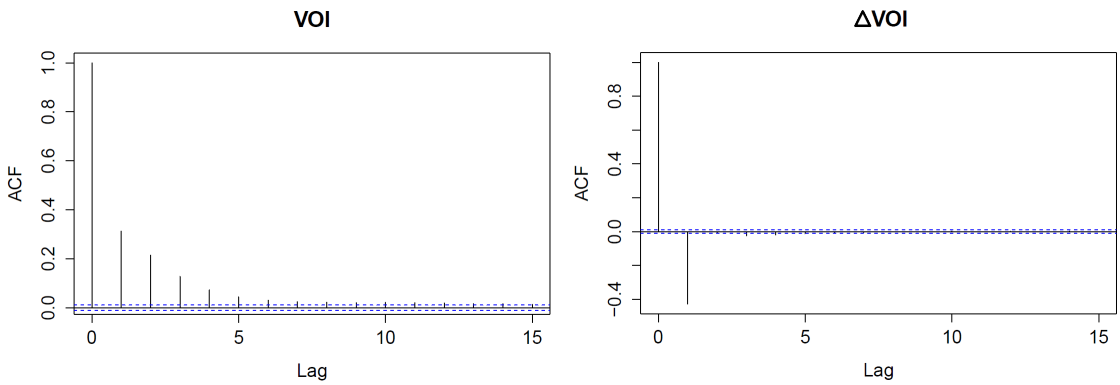

Figure 2.1: Autocorrelation functions for VOI and ΔVOI

We also find that the VOI is positively correlated with contemporaneous price changes. That is, the correlation between OI_t and $\Delta M_t = M_t - M_{t-1}$ is 0.3935 which is consistent with the result in Chordia [3], though the relationship is not as strong. Figure 2.2 below illustrates this positive relationship.

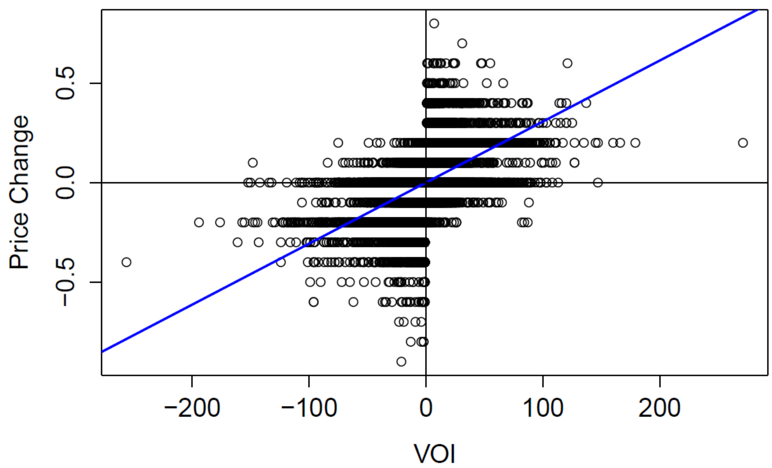

Figure 2.2: Scatterplot of VOI against contemporaneous price change on August 13, 2014. Blue line represents line of best fit.

Furthermore, fitting a contemporaneous linear model $\Delta M_t = \alpha + \beta OI_t + \varepsilon_t$ gives an average daily R^2 of 0.155, which is a much weaker R^2 of 0.69 presented by Cont in [5]. Even if we change the definition of VOI to match the definition of order flow imbalance, it only improves the average R^2 to 0.294. Since both studies are done in the high frequency space, the difference in R^2 it is likely due to the different time scales used in the analysis. By matching the 10 second time interval that Cont uses, we are able to get a daily average R^2 of 0.6537, which is consistent with his results.

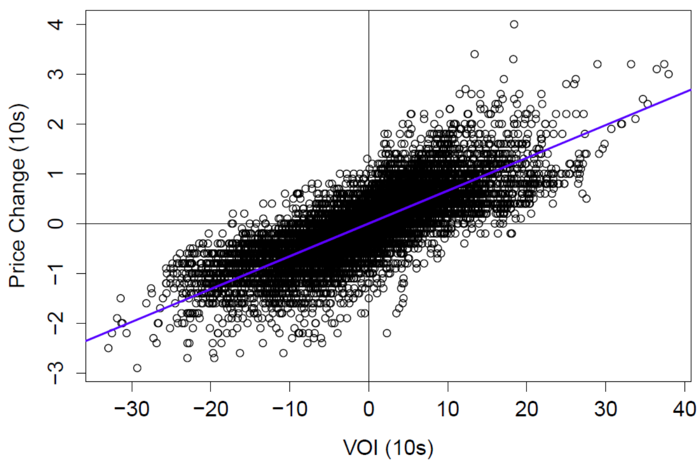

Figure 2.3: Scatterplot of VOI against price change over 10 second intervals on August 13, 2014. Blue line represents line of best fit.

Figure 2.3 above presents the same relationship between VOI and price change at the 10 second interval that is highlighted in [5].

Given that we have found a strong association between VOI and contemporaneous price changes, the following sections will analyze its strength in predicting future price changes and investigate its performance as a trading strategy.

2.2 Assumptions and Setup

We make several assumptions about the trading mechanisms and the simulation of the trading algorithms:

(a) There are no market competitors which means we can always trade at the counterparty price (sell at bid and buy at ask).

(b) There is no latency from the time we receive the new market data to the time we execute the trade, given a favourable signal.

(c) The maximum position allowed is ± 1 contract at any time and we can only buy and sell whole contracts.

(d) The trading cost (commission) is 0.0025% of execution price.

To construct our linear model each day, we select a *main contract* to be traded by examining the trade volume at market open (9:15) by choosing the futures contract with the largest volume. To remove some of the volatility and noise from the data that generally occurs at market open and close [5], we will also be restricting the trading hours from 9:16 to 11:28 and only be allowed to close positions after 11:20 for the morning session, and from 13:01 to 15:13 and only close positions after 15:00 for the afternoon session.

Our strategy uses ordinary least squares to forecast the average mid-price change over the next 10 seconds (20 time-steps) using both instantaneous[1] and lagged VOI. We build the linear model using the previous business day's data and attempt to forecast the mid-price change for the current trading day. This set up is similar to Chordia and Subrahmanyam [4] where they use a linear regression model on lagged order imbalances to predict daily (open-to-close) stock returns. They do not include contemporaneous in the trading strategy as it would be a forward looking measure.

The setup used by Huang *et al.* [7] is slightly different. They calculate order imbalance at 5, 10, and 15 minute intervals in intraday data. After trimming off 90% of small imbalances, the strategy is to buy (sell) when a positive (negative) imbalance

[1] In our case, the instantaneous VOI is calculated the instant the market data is received and hence is *not forward looking*.

appears. Instead of building a linear model to forecast returns, they directly trade on the tails of order imbalance.

2.3 Statistical Analysis

We propose a linear model where we predict future price change with the current (instantaneous) and lagged VOI. Chordia [4], Huang [7], and Cont [5] all build linear models with lagged order imbalances as the explanatory variable and the immediate one-period price change as the response. However for high frequency data, one time-step mid-price changes are often zero. On a daily average, only 45% of one-period price movements are non-zero, so the model would not often capture the larger price movements that the strategy should be trading. Instead, we should consider an average price change over a longer time period, which we will call the *forecast window.* We present some statistical properties of the explanatory and response variables before analyzing the linear model.

The initial linear model suggested by my supervisor, Dr. Zhaodong Wang, is:

$$\overline{\Delta M}_{t,20} = \beta_c + \sum_{j=0}^{5} \beta_j OI_{t-j} + \varepsilon_t \tag{2.3}$$

where $\overline{\Delta M}_{t,20} = \frac{1}{20}\sum_{j=1}^{20} M_{t+j} - M_t$ is the average mid-price change with forecast window $k = 20$, OI_{t-j} is the j-lag VOI, β_c is the constant coefficient, β_j is the j-lag coefficient. We also make the assumption the errors ε_t are independent and normally distributed with zero mean and constant variance (as per Gauss-Markov). This model is constructed independently for each of the 244 trading days in 2014 using ordinary least squares linear regression. Analogous to the study done by Chordia [4], their response variable would be equivalent to setting the forecast window of the average mid-price change in our model to $k = 1$ time-step. From hereon, we will present all linear regression coefficients in a similar style to Chordia [4] and Huang [7].

	Average coefficient	Percent positive	Percent positive and significant*	Percent negative and significant*
Intercept	1.198×10^{-3}	50.41%	29.51%	27.87%
OI_t	2.842×10^{-3}	100.00%	100.00%	0.00%
OI_{t-1}	5.175×10^{-4}	97.95%	88.11%	0.00%
OI_{t-2}	-2.944×10^{-4}	13.52%	3.69%	61.48%
OI_{t-3}	-2.948×10^{-4}	9.43%	2.46%	60.25%
OI_{t-4}	-1.270×10^{-4}	27.46%	7.38%	29.10%
OI_{t-5}	1.569×10^{-5}	50.41%	20.49%	14.34%

* at the 95% confidence level

Table 2.1: Linear Regression Results for Volume Order Imbalance Model. Blue cells indicate majority of the coefficients having significant positive or negative sign.

The average of the daily instantaneous and lagged order imbalance results are shown in Table 2.1 above. We notice that the instantaneous VOI and the lag-1 VOI are both positive and significant at the 5% level for nearly all 244 trading days as indicated by the blue cells. Chordia [3, 4] argues that their lag-1 order imbalance (our instantaneous) has a positive coefficient due to autocorrelated price pressures imposed by past order imbalances. As mentioned in Huang [7], their results do not agree with Chordia's as they found a negative coefficient in their lag-1 order imbalance. One possible reason that both our instantaneous and lag-1 VOI have a positive coefficient is because we take the average price change over 20 time intervals compared to the daily time interval used by Chordia or the 5-minute interval used by Huang. Chordia notes that the remaining lags order imbalance have negative coefficients due overweighting the impact of current imbalances. The price pressures associated with current imbalanced are slowly reversed through time as indicated by negative lag-2 to lag-4 VOI coefficients.

The daily average R^2 of the model (2.3) is 0.0298. This result is consistent with similar research done by Cont [5] on NASDAQ equity data. A plot of the fit for a sample date is shown in Figure 2.4 below. Although the coefficient of the linear model is positive and significant for the instantaneous VOI indicating that there is a positive relationship, it is not immediately obvious in the plot below.

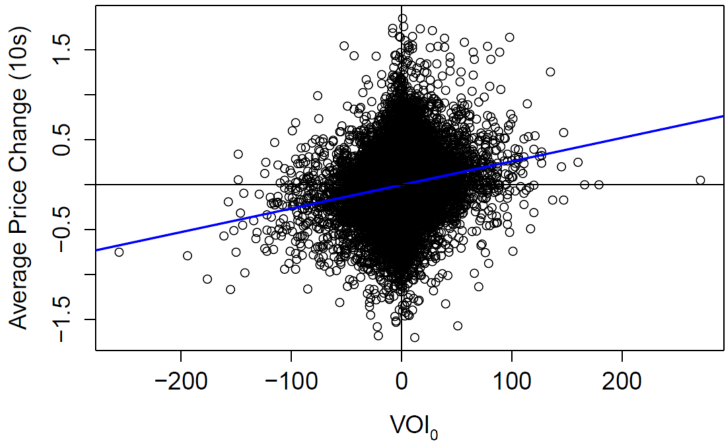

Figure 2.4: Instantaneous VOI vs. Average Price Change on August 13, 2014. Blue line represents line of best fit.

Given that the instantaneous and lagged order imbalance are significantly related with the average 20-step mid-price change, we can use this linear model to devise a trading strategy. At time t, as market data is received, we are able to calculate the instantaneous order imbalance OI_t and obtain a forecast for the mid-price change. If the forecasted price change is greater than 0.2 (less than -0.2) ticks then we buy (sell) the maximum allowed position. The threshold, $q = 0.2$ ticks is chosen as it is the minimum tick size and therefore is also the smallest the bid-ask spread can be. The full trading algorithm is presented in appendix C.

Furthermore, verifying that the price change and VOI process is stationary is important because if there are structural changes in the markets, using previous day's linear model can lead to unfavourable trading signals [10]. Given stationarity, we assume the mean and covariance through time do not change. If the data is not stationary, it would be incorrect to assume that today's price changes can be predicted using previous day's model.

The Augmented Dickey-Fuller (ADF) test and KPSS test are performed on each day's average price change and VOI processes separately. In the case of the ADF test, *not rejecting* the null hypothesis of having a unit root indicates that the process is not stationary. For the KPSS test, rejecting the null hypothesis indicates the process is not stationary. It is important to make the distinction that neither test confirms that the process is stationary, but only if it is *not* not stationary.

ADF Test $H_0 : x_t$ has a unit root $H_1 : x_t$ does not have a unit root			**KPSS Test** $H_0 : x_t$ is stationary $H_1 : x_t$ is not stationary		
	$\overline{\Delta M}_{t,20}$	VOI		$\overline{\Delta M}_{t,20}$	VOI
Reject H_0^*	100.00%	100.00%	Reject H_0^*	8.2%	10.0%
Fail to Reject H_0	0.00%	0.00%	Fail to Reject H_0^*	91.8%	90.0%

* at the 95% confidence level

Table 2.2: Percentage of days the average price change and VOI processes are considered stationary when applying the Augmented Dickey-Fuller and KPSS tests. Favourable results are highlighted in green.

Table 2.2 above summarizes the tests of stationarity and reports the proportion of dates whose average price change and VOI processes that pass each test at the 95% confidence level. Using the ADF test, we see that the average price change and VOI processes both reject the null hypothesis of having a unit root. Further supporting the idea that our response and explanatory variables are stationary is the results from the KPSS test – no more than 10% of the days reject the null hypothesis of stationarity. This means on majority of trading days, we should be able to apply the same trading strategy to generate positively increasing profits due to the stationarity of our time-series processes [14].

2.4 Results and Performance

The results of using the linear model (2.3) as a trading strategy is summarized in the table below. The full set of results for this strategy is in appendix A.1.

Statistical test: one-tailed, one-sample t-test (df = 243) $H_0 : \mu \geq 0$ $H_1 : \mu < 0$						
Mean daily profit (CNY)	Standard error	t-stat	p-value	Days with profit	Days with losses	Mean daily trade volume
19,528	3,290	5.9352	5.032×10^{-9}	185	58	634
Average Daily Sharpe Ratio: 0.380 Annualized Sharpe Ratio: 5.935						

Table 2.3: Trading strategy results using volume order imbalance linear model

The results in Table 2.3 indicate that the strategy yields a statistically significant positive average daily profit and approximately 76% of the trading days generates positive returns. If we assume a margin of 100,000 CNY, the mean daily return is approximately 19.5%. Despite order imbalance having a low correlation with the average price change, the strategy performs quite well. One reason for this is due to using the predicted price change as a multinomial classifier used for decision-making rather than for additional calculations. On average, the daily correlation between the actual and predicted price change is only 0.166 but if we transform the prediction into a three-class variable $\{-1, 0, +1\}$ based on the trading threshold ± 0.2, the correlation increases to 0.449.

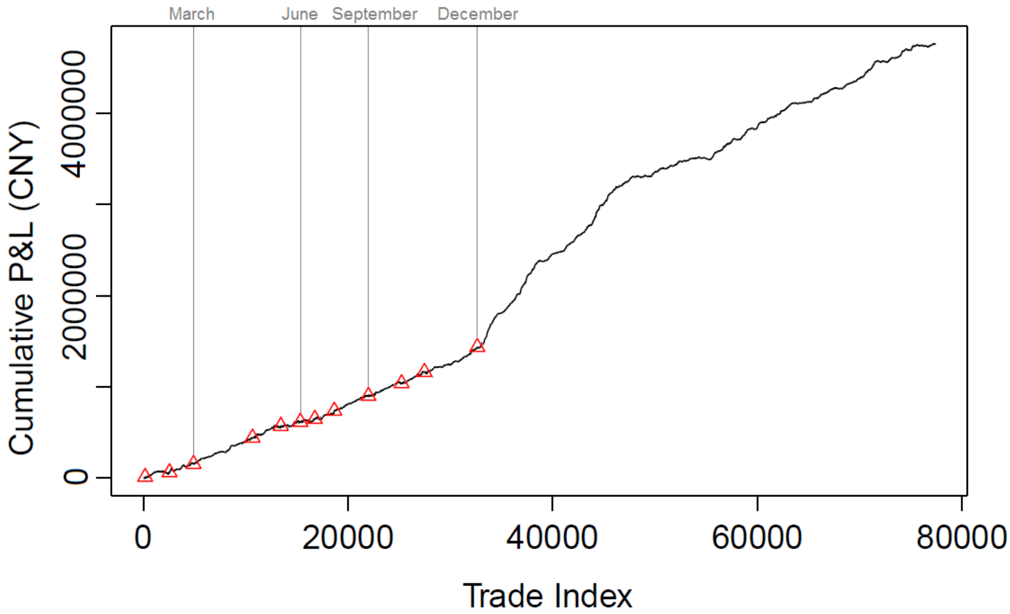

Figure 2.5: Cumulative P&L by trade; red points indicate beginning of month

In Figure 2.5 above, we notice there is a linear relationship between profit and the number of trades and also time in the first 11 months of trading. Each red point in the plot represents the beginning of a month. The cumulative P&L after December grows at a greater rate then the previous 11 months and more than half of the trades were made in December alone. So what happened in December?

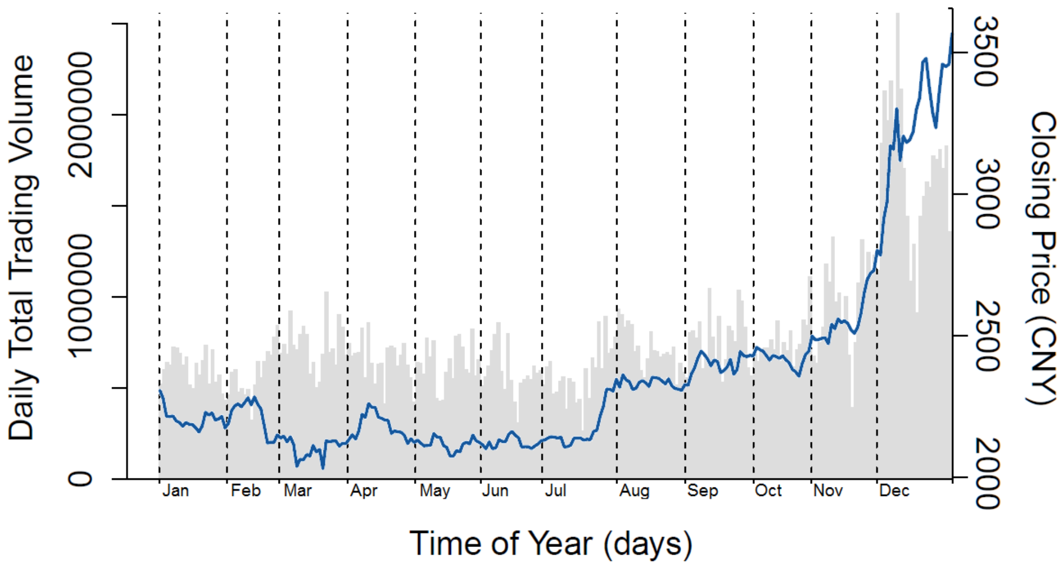

Figure 2.6: Daily transactional trading volume for main contract by day in grey (left-axis) overlayed by daily close mid-price in CNY in blue (right-axis) with each dotted line indicating beginning of a new month

From Figure 2.6, we can see that the daily traded volume of the main futures contract nearly doubled in December compared to the previous 11 months. This is likely due to the price of the main contract rising at the same time. As the price rises, trading the asset becomes more attractive to more investors so volumes rise with it. This is evident based on the highly positive correlation between total traded daily volume and daily close price (0.828). In addition to higher prices and higher volumes, our strategy trades, on average per day, 13 times more and generates nearly 25 times as much profit in December compared to the rest of the year. The mean daily profit is highly positively correlated with total traded daily volume (0.863). There is a clear change in market conditions in December and we will address the strategy's potential shortcomings in the next section.

Chordia [4] reported a statistically significant positive average daily return on their order imbalanced strategy of approximately 0.09%. However, after accounting for potential transaction costs and commissions, he reasoned that the profits would be nullified. The strategy employed by Huang [7] generated negative returns when trading on counterparty price but managed to give significant positive returns (0.49%) when trading on transaction price (mid-price, essentially).

Another study done by Ravi and Sha [11] on equity markets from 1993 to 2010 finds that a strategy based on buying and selling pressure also produces significant positive returns. Similar to Chordia and Huang, they use Lee and Ready's algorithm

[9] to classify the directional pressure of each stock. Instead of using a linear model (as in Chordia [4]) or a percentile threshold signal (as in Huang [7]), they rank the top 10% of stocks with greatest buy (sell) pressure and form a buy (short) portfolio to trade each day. This method produced returns ranging from 7.3 basis points (in 2007) to 47 basis points per day (in 1993).

The results of each study indicate that order imbalance can be used to yield statistically significant positive returns and is consistent with our findings when the strategy is applied at a high frequency level. Most notably, the strategy performs extremely well when applied to high frequency Chinese Index Futures and even better when volume is abnormally high (as it was in December 2014). This strategy heavily relies on choosing a futures contract with high trading volume thereby being able to produce a strong imbalance signal which is derived from bid-ask volume.

Lastly, we find that the forecast window $k = 20$ is not optimal for this linear model and can be parametrized. Figure 2.7 below shows, for fixed lag, the daily mean profit for various forecast windows $k = 1, ..., 30$. We see that it peaks at $k = 4$ before decreasing linearly in k and gives us an idea of the optimal forecast window for the parametrized model.

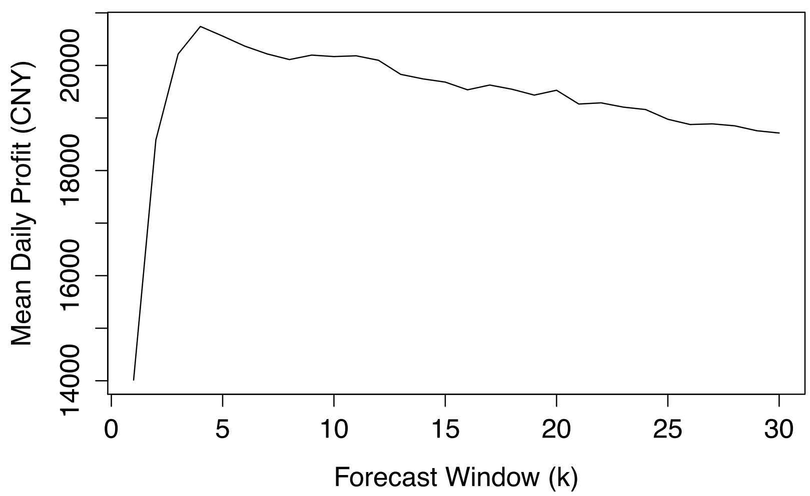

Figure 2.7: Mean daily profit for various forecast windows

A more detailed analysis on how average daily profits change with forecast window is presented in the next chapter.

2.5 Summary and Considerations

The order imbalance strategy proposed by Dr. Wang is highly successful on the Chinese Index Futures market when applied to high frequency data. Similar strategies designed by Chordia [4], Huang [7], and Ravi [11] all produce positive average daily returns when applied to larger time-scales, although the magnitude of these returns are nearly 200 times smaller than ours. Cont [5] also presented a similar positive relationship between order flow imbalance and future price change in high frequency NASDAQ data.

As mentioned in the previous section, daily profit is highly correlated with daily trade volume so applying this strategy to futures with lower volumes would likely perform poorly. If there's not enough volume, the bid-ask spread might be very wide and it would not be beneficial to trade at counterparty price. When applying the strategy on a futures contract with the second highest trading volume at market open, the mean daily profit is only 2,801 CNY (down 86% from main contract profit) and the win ratio is down nearly 30% from 76% to 49%. This means on more than half of the trading days, this strategy loses money trading a futures contract with lower volume. Finally, the annualized Sharpe ratio falls from 5.935 to 1.763. The full profit and loss results for applying the strategy to the secondary contract can be found in appendix A.2

Lastly, the linear model (2.3) considers only a single factor: the VOI up to lag 5. Previous sections showed that when fitting a linear model using ordinary least squares, the coefficients are statistically significant. Despite performing well as a trading strategy, the fit of the linear model was quite poor and does not explain the variation in price change well. However, we find that if we treated the predicted price change as a multinomial class variable, its correlation with the actual price change improved from 0.166 to 0.449. This suggests that we can improve our prediction if we considered additional features in the linear model. One drawback to the VOI is that it only considers the size of the order imbalance but not degree or strength of the imbalance. The next chapter will elaborate on additional factors and how they can capture more detailed imbalance information in the high frequency dataset.

Chapter 3

Improved Strategy

We would like to improve upon the existing strategy presented in chapter 2 by extending the linear model with new factors and also by optimizing the regression and trading parameters. The improvements described in the following section produces a net daily profit of 58,600 CNY, a win ratio of 94.7%, and daily Sharpe ratio of 0.464. This is a 400% increase in profit, nearly 20% increase in dates with positive profits, and 22% increase in Sharpe ratio.

3.1 Additional Factors and Analysis

3.1.1 Order Imbalance Ratio

The VOI only measures the magnitude of the imbalance which is not sufficient to describe the behaviour of the traders in the market. For example, if the current bid change volume is 300 and the current ask change volume is 200, the VOI is 100, which is considered a strong signal to buy. However, this does not take into consideration the ratio between the bid volume and ask volume which indicates the strength of the potential buyers in the market. Hence, we define a new factor called the Order Imbalance Ratio (OIR), as:

$$\rho_t = \frac{V_t^B - V_t^A}{V_t^B + V_t^A} \tag{3.1}$$

This factor complements the volume order imbalance by allowing us to distinguish cases where the difference is large but the ratio is small. In the example presented above, the OIR is only 0.2, indicating that the original signal to buy may not be that strong after all.

The OIR is another measure of order imbalance and should share similar statistical properties to VOI. The autocorrelation is presented below in Figure 3.1 and they share the same signs and similar magnitudes with the autocorrelation of VOI in Figure 2.1.

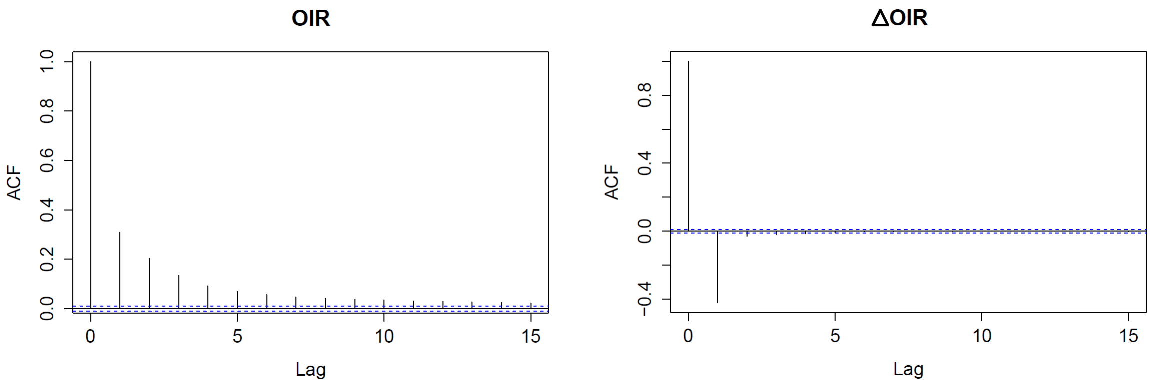

Figure 3.1: Autocorrelation functions for OIR and ΔOIR

However, we find that the relationship between OIR and contemporaneous price change is actually the opposite to VOI. The correlation between ρ_t and ΔM_t is -0.3458. One interpretation of this is due to the same order-splitting theory by Chordia [4]. Since the autocorrelations of ρ_t is also significant and positive for the first 5 lags as VOI was, we can assume this is an equivalent representation of the order-splitting behaviour of traders. By definition, a large OIR means that the bid volume is much greater than the ask volume at a given time, indicating many traders have the intention to buy, and very few have the intention to sell. But since a large OIR is associated with a negative price change, it means that more a larger *proportion* of traders are willing to buy when prices have fallen. This result demonstrate *how* the orders are split over time as opposed to the autocorrelations which only indicates the presence of order-splitting.

3.1.2 Mean Reversion of Mid-Price

Aside from the VOI and OIR, we include a way to classify trades as being buyer-initiated or seller-initiated. Using the traded volume and turnover information in the data set, we are able to determine the average trade price between two time-steps. We define the Average Trade Price, $\overline{TP}_t$ from $(t-1, t]$ as:

$$\overline{TP}_t = \begin{cases} M_1, & t = 1 \\ \frac{1}{300}\frac{T_t - T_{t-1}}{V_t - V_{t-1}}, & V_t \neq V_{t-1} \\ \overline{TP}_{t-1}, & V_t = V_{t-1} \end{cases} \tag{3.2}$$

where T_t is the turnover (trade volume in CNY) and V_t is the transaction volume at time t. This process represents the average price that other market participants executed their trades at which can be interpreted as a proxy for trade imbalance. By checking whether $\overline{TP}$ is closer to the ask (bid) price, we can classify trades as being more buyer (seller) initiated. However, instead of a binary classification, we define the factor as the distance of the average trade price from the average mid-price over the time-step $(t-1, t]$:

$$R_t = \overline{TP} - \frac{M_{t-1} + M_t}{2} = \overline{TP}_t - \overline{MP}_t \tag{3.3}$$

where M_t is the mid-price at time t. The factor R_t, which we call the *mid-price basis* (MPB), is an important predictor of price change because of its mean reversive properties. It gives a continuous classification of whether trades were buyer or seller initiated. A large positive (negative) quantity means the trades were, on average, closer to the ask (bid) price. Our definition of the MPB is similar to the definition of order imbalance definition used in [3, 4, 11] where they use the Lee and Ready algorithm [9] to classify trades as buyer or seller initiated.

To check whether the process R_t is mean-reverting, we apply the variance ratio (VR) test outlined in [2]. For the time series $R_t = \phi R_{t-1} + \varepsilon_t$, the null hypothesis of the VR test is $H_0 : \phi = 1$. If the series is a random walk ($\phi = 1$), then for a k-period lag, we get the relationship:

$$\begin{aligned} R_t - R_{t-k} &= (R_t - R_{t-1}) + (R_{t-1} - R_{t-2}) + \ldots + (R_{t-k+1} - R_{t-k}) \\ &= \varepsilon_t + \varepsilon_{t-1} + \ldots + \varepsilon_{t-k+1} \\ &= \sum_{j=0}^{k-1} \varepsilon_{t-j} \end{aligned} \tag{3.4}$$

Assuming that the errors are independent and identically normally distributed[1] with zero mean and constant variance, taking the variance of both sides gives:

$$\begin{aligned} \hat{\sigma}^2(k) = \mathbb{V}\big[R_t - R_{t-k}\big] &= \mathbb{V}\big[\sum_{j=0}^{k-1} \varepsilon_{t-j}\big] \\ &= k\mathbb{V}\big[\varepsilon_t\big] \\ &= k\mathbb{V}\big[R_t - R_{t-1}\big] = k\hat{\sigma}^2(1) \end{aligned} \tag{3.5}$$

Hence, for a process which follows a random walk, we test whether variance ratio $\frac{\hat{\sigma}^2(k)}{k\hat{\sigma}^2(1)}$ is equal to 1. In this paper, we will not address the issue of a finite sample size as our daily sample size (around 30,000) is likely considered large enough for a variance ratio test of up to lag 20. Figure 3.2 below depicts the ratios for $k = 2$ to

[1]For most dates, both the ADF-test and KPSS-test indicate R_t is a stationary process.

100 and indicate that they are below 1 for all lags, meaning the process R_t exhibits mean-reversion [2]. The result is statistically significant at the 99% confidence level.

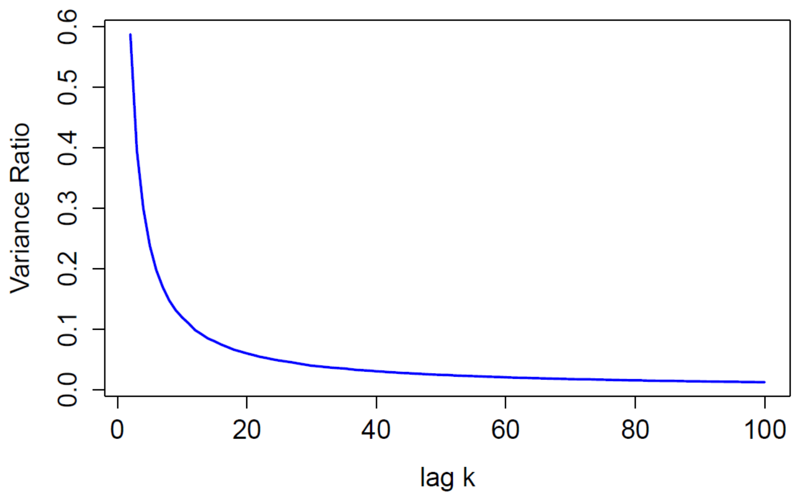

Figure 3.2: Variance ratios for R_t on August 13, 2014

We expect that R_t will revert back to mean 0 so if $R_t > 0$, the mid-price will eventually increase and revert towards the average trade price and if $R_t < 0$, then we would expect the mid-price to decrease back to average trade price. Thus, we have a buy signal when $R_t > 0$ and sell signal when $R_t < 0$. The positive relationship between R_t (MPB) and the average mid-price change (response) is shown below in Figure 3.3.

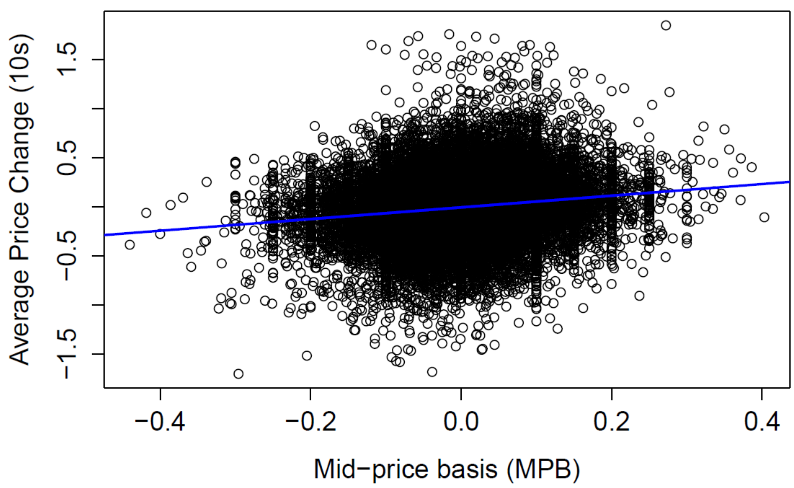

Figure 3.3: Scatterplot of the mid-price basis (R_t process) against the response variable (average price change) on August 13, 2014. Blue line indicates line of best fit.

3.1.3 Bid-Ask Spread

The bid-ask spread at time t is defined as $S_t = P_t^A - P_t^B$. It is an important measure of liquidity and has a positive relationship with contemporaneous price volatility and negative relationship with trade volume based on findings by Wang and Yau [13] for various CME Futures. This information can be used to adjust our regression factors for different levels of liquidity by dividing them by instantaneous spread. This spread adjustment was found in collaboration with Xuan Liu [10].

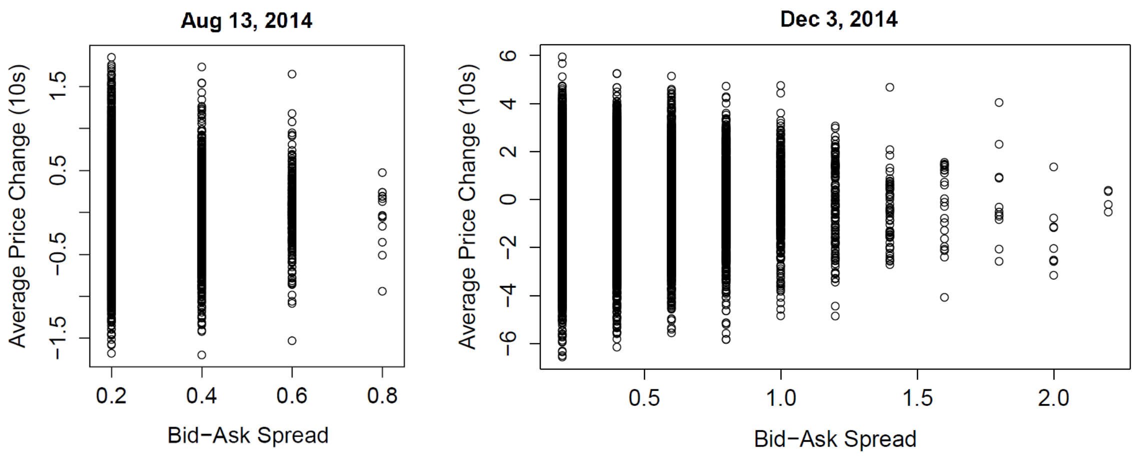

Figure 3.4: Scatterplot of spread against average price change for August 13, 2014 (left) and December 3, 2014 (right) illustrating high spreads are associated for price changes near zero.

From Figure 3.4 above, we see that when spread is large, there are very few observations where the average change in mid-price is away from zero. Hence when liquidity is low, the price is slow to change and therefore trading may be unfavourable. This will potentially reduce the risk of trading on a weak signal when the spread is large.

One interesting observation is that a large VOI (negative or positive) is associated with smaller spreads. From Figure 3.5 below, we notice that large spreads only occur when order imbalance is near 0. Chordia [3] finds that higher spreads are associated with a larger (absolute) order imbalance but our findings above does not agree with his results. This may be due to the fact that we are using high frequency data and not aggregated daily data.

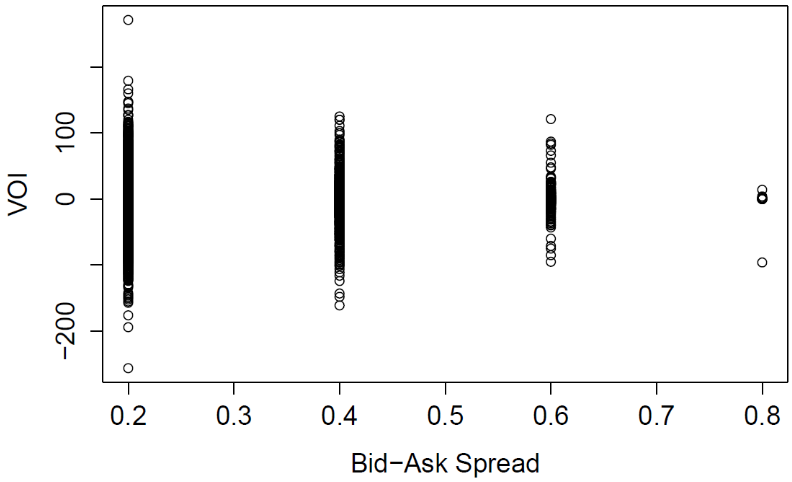

Figure 3.5: Scatterplot of spread against VOI for August 13, 2014 illustrating high spreads are associated with VOI near zero.

Using these relationships between spread and price change and spread and order imbalance, we adjust our factors so that we would not falsely obtain a trading signal when spread is large. The final parametrized model which includes the spread adjustment is presented in the following section.

3.2 Parameter Selection and Results

3.2.1 Parametrized Linear Model

We incorporate the new features OIR and MPB as defined by (3.1) and (3.3) respectively into our linear model. Each feature will also include the spread adjustment discussed in 3.1.3 by dividing by the spread. The final linear model is presented in equation (3.6) below.

$$\overline{\Delta M}_{t,k} = \beta_0 + \sum_{j=0}^{L} \beta_{OI,j} \frac{OI_{t-j}}{S_t} + \sum_{j=0}^{L} \beta_{\rho,j} \frac{\rho_{t-j}}{S_t} + \beta_R \frac{R_t}{S_t} + \varepsilon_t \tag{3.6}$$

where $\overline{\Delta M}_{t,k} = \frac{1}{h} \sum_{j=1}^{k} M_{t+j} - M_t$ is the k-step average mid-price change, OI_{t-j} is the j-lag Volume Order Imbalance from the previous strategy, ρ_{t-j} is the j-lag Order Imbalance Ratio, R_t is the instantaneous mid-price basis, and S_t is the instantaneous bid-ask spread. We also parametrize the lag L for Volume Order Imbalance and Order Imbalance Ratio. For coefficients, β_0 is the constant term, $\beta_{OI,j}$ corresponds to the j-lag spread-adjusted VOI, $\beta_{\rho,j}$ corresponds to the j-lag spread-adjusted OIR, and β_R corresponds to the spread-adjusted MPB. The errors ε_t are assumed to be independent and identically normally distributed with zero mean and constant variance. This model will be built using ordinary least squares linear regression.

We note from the previous model (2.3) that the parameters are set to $k = 20$ and $L = 5$ even though they may not be optimal for this strategy. The next section will discuss parameter optimization.

3.2.2 Comparison with Order Imbalance Strategy

Without any parameter optimization, we can compare the results of the two linear models (2.3) and (3.6) given the same forecast window $k = 20$, lags for order imbalance $L = 5$ (and trading threshold $q = 0.2$):

$$\overline{\Delta M}_{t,20} = \beta_0 + \sum_{j=0}^{5} \beta_{OI,j} \frac{OI_{t-j}}{S_t} + \sum_{j=0}^{5} \beta_{\rho,j} \frac{\rho_{t-j}}{S_t} + \beta_R \frac{R_t}{S_t} + \varepsilon_t \tag{3.7}$$

That is, we will use the previous day's linear model to forecast today's 20-step average mid-price change and only trade if the change is above 0.2 (below -0.2). The linear model (3.7) has $R^2 = 0.0701$ which is greater than the previous model (2.3).

	Average coefficient	Percent positive	Percent positive and significant	Percent negative and significant
Intercept	1.131×10^{-3}	50.00%	31.97%	30.33%
$OI_t S_t^{-1}$	4.992×10^{-4}	100.00%	100.00%	0.00%
$OI_{t-1} S_t^{-1}$	1.879×10^{-4}	100.00%	100.00%	0.00%
$OI_{t-2} S_t^{-1}$	-3.698×10^{-5}	36.89%	13.52%	31.97%
$OI_{t-3} S_t^{-1}$	-6.894×10^{-5}	11.89%	4.51%	55.74%
$OI_{t-4} S_t^{-1}$	-4.224×10^{-5}	21.31%	6.15%	38.93%
$OI_{t-5} S_t^{-1}$	-7.047×10^{-6}	39.75%	15.16%	18.44%
$\rho_t S_t^{-1}$	1.610×10^{-2}	99.59%	92.62%	0.00%
$\rho_{t-1} S_t^{-1}$	-1.147×10^{-2}	0.00%	0.00%	100.00%
$\rho_{t-2} S_t^{-1}$	-1.349×10^{-3}	4.51%	0.00%	52.46%
$\rho_{t-3} S_t^{-1}$	1.148×10^{-3}	74.59%	20.90%	0.82%
$\rho_{t-4} S_t^{-1}$	1.240×10^{-3}	82.79%	29.92%	0.00%
$\rho_{t-5} S_t^{-1}$	1.214×10^{-3}	82.38%	32.38%	1.64%
$R_t S_t^{-1}$	1.038×10^{-1}	100.00%	100.00%	0.00%

Table 3.1: Linear regression results for Improved Model (3.7) using the same parameters as the Volume Order Imbalance model (2.3): $k = 20, q = 0.2, L = 5$. Blue cells indicate majority of the coefficients having a significant positive or negative sign.

Similar to the previous results in Table 2.1, the instantaneous and lag-1 VOI remain significant positive factors as indicated by the blue cells. Furthermore, the new order imbalance ratio factors are also significant at least half the time up to lag-2. More interestingly, the instantaneous order imbalance ratio has a positive relationship with price change while the lag-1 and lag-2 OIR have a negative relationship. Using this definition of order imbalance, these coefficients are consistent with the findings in Chordia [4]. Evidence of the price pressure (induced by the instantaneous OIR) reversal by the lag-1 OIR coefficient is very pronounced as they were negative and significant for every trading day.

This improved linear model gives a mean daily profit of 50,369 CNY and a win-ratio of 92.6% compared with the previous volume order imbalance strategy with mean daily profit 19,528 CNY and win-ratio of 75.8%. A one-tailed paired t-test ($H_0 : \mu_{new} - \mu_{old} \leq 0, H_1 : \mu_{new} - \mu_{old} > 0$) on the daily net profit gives a p-value of 1.2317×10^{-9}, a highly significant result, and thus rejecting the null hypothesis.

Simply by adding new factors to the linear model to represent the price pressure and trader intentions improves the mean daily profit by over 350%. We find that volume order imbalance alone is an inadequate measure of the buying and selling pressures in the market. By including the OIR and MPB factors, we have a better understanding of trader intention as they submit orders to the market.

3.2.3 Parameter Analysis

We consider several optimizations in both the linear regression parameters and the trading parameters. The regression coefficients for the linear models exhibit strong daily autocorrelation as seen in Figure 3.6 below.

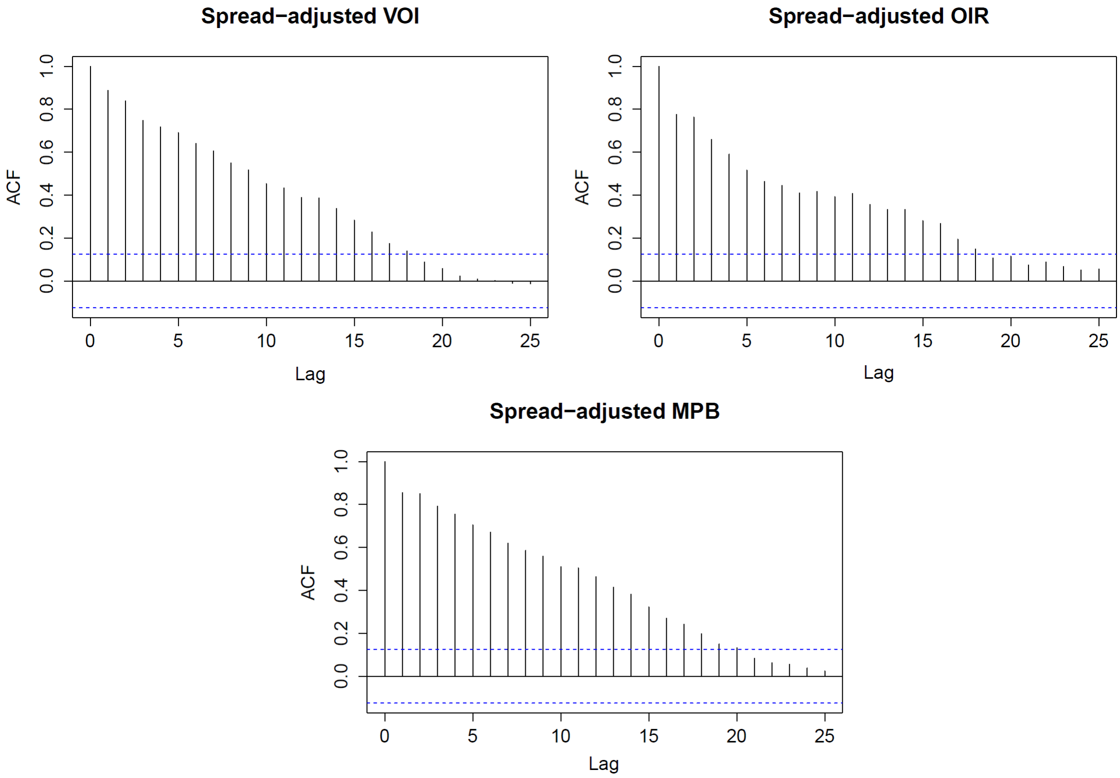

Figure 3.6: ACF of spread-adjusted VOI (top-left), spread-adjusted OIR (top-right) and spread-adjusted MPB (bottom). The plots indicate the factor coefficients (the daily β_is) exhibit strong daily autocorrelation.

Instead of simply using the previous day's regression coefficients, consider using a weighted moving average of the coefficients from the past p days:

$$\hat{\beta}_i^{(d)} = \sum_{j=1}^{p} w_j \beta_i^{(d-j)}, \qquad \sum_{j=1}^{p} w_j = 1$$

One method to estimate the weights w_j is to fit an AR(p) model to the most significant regression coefficients ($OI_t S_t^{-1}$, $\rho_t S_t^{-1}$, and $R_t S_t^{-1}$) and set them proportional to the AR coefficients. Another method would be to simply take the simple moving average of the past p days.

We test the strategy for fixed parameters $k = 20$ and $L = 5$ using various weights for the lagged coefficients. Test results in Table 3.2 show that the 2-day simple moving average, $w_1 = w_2 = 0.5$, performs better than having no weighted average at all, the 3-day and 4-day simple moving averages. Using weights proportional to the spread-adjusted VOI, OIR, or MPB AR(2) models does not perform any better than the 2-day simple moving average. For simplicity, we will proceed with the 2-day simple average model for the rest of this chapter.

Statistical tests: one-tailed paired t-test (df = 243) $H_0 : \mu_2 - \mu_j \leq 0$ $H_1 : \mu_2 - \mu_j > 0$							
	2-day simple	3-day simple	4-day simple	OI_t-AR(2) weighted	ρ_t-AR(2) weighted	R_t-AR(2) weighted	previous day only
mean daily P&L*	51,713	50,749	50,753	51,305	51,730	51,721	50,369
t-stat	–	2.2694	1.6865	0.9974	-0.1631	-0.0962	1.7927
p-value	–	0.01206	0.04649	0.1598	0.5647	0.5383	0.03713

* in CNY

Table 3.2: Strategy results for various lagged coefficient weights and parameters $k = 20, q = 0.2, L = 5$. The AR(2) weights for: OI_t are $(0.776, 0.224)$, ρ_t are $(0.531, 0.469)$, and R_t are $(0.516, 0.484)$. The 2-day simple moving average performs better or just as well as the other coefficient weights.

We also test the strategy using various lags for the spread-adjusted VOI and OIR variables. Since lag selection is essentially feature selection for the parametrized linear model (3.6), we attempt to choose the model with the best fit using a step-wise algorithm. The goodness-of-fit will be measured by the Akaike information criterion (AIC) and results are presented in Table 3.3 below. Additonally, the results of one-tailed paired t-tests on daily profits for the various lags $L = 0, 1, ..., 7$ are also shown.

Statistical tests: one-tailed paired t-test (df = 243) $H_0 : \mu_{L=5} - \mu_{L=j} \leq 0$ $H_1 : \mu_{L=5} - \mu_{L=j} > 0$								
	lag-5	no-lag	lag-1	lag-2	lag-3	lag-4	lag-6	lag-7
mean daily P&L*	51,713	43,811	49,457	50,264	51,091	51,190	51,740	51,642
t-stat	–	6.2299	4.4429	3.9286	2.0689	2.2251	-0.14	0.3321
p-value	–	< 0.01	< 0.01	< 0.01	0.01981	0.0135	0.5556	0.37
AIC	19,931	20,375	20,099	20,045	19,998	19,961	19,905	19,880

* in CNY

Table 3.3: Strategy results for lags $L = 0, 1, ..., 7$ using 2-day simple moving average for coefficient weights and parameters $k = 20, q = 0.2$. Notice that $L = 5$ outperforms $L = 0, .., 4$ and is no different than $L = 6$ or 7 at the 95% confidence level. Lag-7 model has lowest AIC, highlighted in green.

We see that the lag selection has a significant effect on total profit when the coefficient weights, forecast window, and trading threshold are held fixed. Lag $L = 5$ performs better than the no-lag, lag-1, and lag-2 models at the 99% significance level; better than the lag-3 and lag-4 models at the 95% significance level; and no different than the lag-6 and lag-7 models. Furthermore, AIC indicates that the lag-7 linear model has the best fit despite having slightly lower P&L than the lag-5 model (although insignificant). This suggests that a model's goodness-of-fit is not necessarily the best indicator of strategy performance. Out of model simplicity, we will keep $L = 5$ as the optimal lag for VOI and OIR variables.

3.2.4 Parameter Selection Results

Given the previous analysis done for the coefficient lag and the variable lag, we can now focus on selecting the optimal forecast window and trading threshold. Optimal parameter selection is done by running the strategy on a mesh defined by $q \in [0.13, 0.20]$, $\Delta q = 0.005$ and $k = 1, 2, ..., 20$. We will choose the optimal k and q with the greatest mean daily profit and calculate 90%, 95%, and 99% confidence regions for the parameters by inferring directly from the confidence regions of the daily profit. The reason we choose $q = 0.20$ as the upper bound is because increasing the threshold will only result in less trades and lower overall profit.

After fitting the linear model (3.6) with lag $L = 5$ over different forecast windows, we can check how well the new factors, spread-adjusted VOI, OIR, and MPB, can explain the variation in the price change by comparing the R^2. However, as mentioned in the previous chapter, we should be careful looking at the goodness-of-fit of the models since a better fit does not necessarily indicate a more profitable strategy. From Figure 3.7 below, the R^2 peaks when $k = 2$ and decreases as the forecast window increases. We will check the performance of the strategy for various forecast windows and see whether $k = 2$ actually generates the most profit.

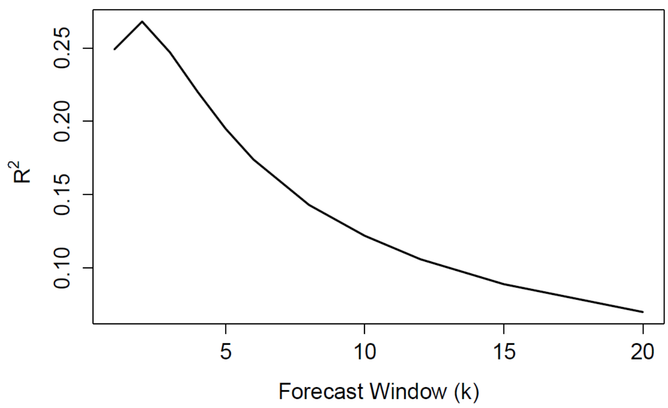

Figure 3.7: R^2 for model (3.6) for forecast windows $k = 1$ to 20.

We present a heatmap in Figure 3.8 summarizing the profits on the mesh defined by $k = 1, ..., 20$ $q \in [0.13, 0.20]$, $\Delta q = 0.005$ on the following page. Darker blue cells indicate a larger average daily P&L while darker red cells denote lower P&L.

Trading Threshold (q)

Forecast Window (k)	0.13	0.135	0.14	0.145	0.15	0.155	0.16	0.165	0.17	0.175	0.18	0.185	0.19	0.195	0.2
1	47,875	47,992	46,984	45,827	44,661	43,498	42,879	42,176	41,099	39,969	38,896	38,199	37,382	36,695	35,451
2	54,510	55,621	55,413	54,801	54,300	52,927	52,057	50,778	49,502	48,550	47,770	46,443	45,593	44,249	43,270
3	56,031	57,227	57,827	57,862	57,163	56,623	55,419	54,596	53,449	52,013	50,735	49,904	49,095	48,139	47,238
4	56,378	57,435	58,059	58,299	58,313	57,417	56,695	56,068	54,968	53,980	52,746	51,499	50,475	50,018	48,946
5	55,798	57,089	58,099	58,408	$\boxed{58,600}$	58,162	57,032	56,382	55,692	54,681	53,705	52,606	51,402	50,766	49,692
6	55,792	57,128	57,773	58,220	58,360	58,248	57,244	56,554	55,855	54,931	54,138	53,089	52,252	51,199	50,391
7	55,274	57,104	57,860	58,201	58,344	58,356	57,309	56,847	55,921	54,957	54,287	53,333	52,427	51,335	50,775
8	55,110	57,021	57,769	58,243	58,224	58,348	57,780	57,102	55,904	55,117	54,166	53,504	52,883	51,638	50,705
9	54,885	56,804	57,694	58,125	58,327	58,122	57,863	57,279	56,195	55,159	54,385	53,497	52,920	51,876	50,646
10	54,664	56,643	57,721	58,024	58,322	58,157	57,883	57,208	56,187	55,347	54,703	53,664	53,042	52,065	50,867
11	54,528	56,548	57,685	57,896	58,304	58,068	57,815	57,255	56,349	55,393	54,951	53,884	53,061	52,210	51,049
12	54,224	56,304	57,496	57,959	58,145	57,979	57,720	57,360	56,608	55,571	55,000	53,984	53,105	52,290	51,194
13	53,838	56,212	57,284	57,891	58,077	57,928	57,715	57,307	56,524	55,768	55,031	54,189	53,271	52,270	51,240
14	53,745	56,054	57,273	57,586	58,090	57,862	57,825	57,305	56,659	55,753	55,141	54,192	53,308	52,220	51,256
15	53,670	55,857	57,100	57,676	57,984	57,972	57,803	57,242	56,781	55,726	55,255	54,351	53,410	52,241	51,365
16	53,484	55,688	56,798	57,548	57,771	57,931	57,739	57,310	56,825	55,914	55,315	54,288	53,301	52,328	51,400
17	53,219	55,514	56,715	57,316	57,711	58,023	57,699	57,259	56,950	56,012	55,420	54,333	53,259	52,418	51,375
18	53,091	55,199	56,534	57,130	57,650	58,015	57,654	57,248	56,998	56,134	55,398	54,337	53,255	52,444	51,502
19	52,972	55,003	56,429	57,079	57,541	57,864	57,691	57,313	56,933	56,196	55,395	54,369	53,160	52,512	51,528
20	52,897	54,760	56,176	57,056	57,515	57,744	57,694	57,283	57,105	56,165	55,392	54,546	53,308	52,464	51,713

Figure 3.8: Mean daily profit and loss heatmap for the strategy using the linear model (3.6) with lag $L = 5$ and coefficient weights $w_1 = w_2 = 0.5$ over the mesh $q = 0.13, 0.135, ..., 0.2$ and $k = 1, ..., 20$. Darker blue cells denote larger P&L while darker red cells denote lower P&L. The parameters with the largest mean daily P&L is $k = 5$ and $q = 0.15$ indicated by the thick border.

From Figure 3.8 above, we notice that the parameters with the largest mean daily profit is the pair $(k, q) = (5, 0.15)$ as indicated by the thick border. The linear model constructed on day d, associated with these parameters is:

$$\overline{\Delta M}_{t,5} = \beta_0^d + \sum_{j=0}^{5} \beta_{OI,j}^d \frac{OI_{t-j}}{S_t} + \sum_{j=0}^{5} \beta_{\rho,j}^d \frac{\rho_{t-j}}{S_t} + \beta_R^d \frac{R_t}{S_t} + \varepsilon_t \tag{3.8}$$

Revisiting the goodness-of-fit for the models in Figure 3.7, we notice that despite having a lower R^2, a forecast window of 5 performs better than the model with a higher R^2 (such as $k = 2$). One possibility is that the linear models for $k = 1$, 2, or 3 may fit well for the current day but is not a good model for the following business days to be used as a trading strategy.

	Average coefficient	Percent positive	Percent positive and significant	Percent negative and significant
Intercept	3.088×10^{-4}	51.64%	19.67%	15.57%
$OI_t S_t^{-1}$	4.458×10^{-4}	100.00%	93.44%	0.00%
$OI_{t-1} S_t^{-1}$	1.868×10^{-4}	100.00%	100.00%	0.00%
$OI_{t-2} S_t^{-1}$	-2.452×10^{-5}	43.03%	20.08%	33.61%
$OI_{t-3} S_t^{-1}$	-6.520×10^{-5}	7.38%	2.87%	79.92%
$OI_{t-4} S_t^{-1}$	-4.553×10^{-5}	12.30%	1.64%	68.44%
$OI_{t-5} S_t^{-1}$	-1.625×10^{-5}	28.28%	7.38%	38.11%
$\rho_t S_t^{-1}$	1.876×10^{-2}	100.00%	100.00%	0.00%
$\rho_{t-1} S_t^{-1}$	-1.026×10^{-2}	0.00%	0.00%	100.00%
$\rho_{t-2} S_t^{-1}$	-1.745×10^{-3}	0.41%	0.00%	92.21%
$\rho_{t-3} S_t^{-1}$	6.243×10^{-4}	57.79%	19.67%	5.74%
$\rho_{t-4} S_t^{-1}$	7.960×10^{-4}	76.23%	32.38%	1.23%
$\rho_{t-5} S_t^{-1}$	7.879×10^{-4}	76.23%	33.61%	2.87%
$R_t S_t^{-1}$	8.228×10^{-2}	100.00%	99.18%	0.00%

Table 3.4: Linear Regression Results for Improved Model (3.8) using the optimal parameters: $k = 5, q = 0.15, L = 5, w_1 = 1$. Blue cells indicate majority of the coefficients having a significant positive or negative sign.

Table 3.4 above shows the average coefficients for model (3.8) with $k = 5$, which are comparable to the coefficients presented in Table 3.1 for model (3.7) with $k = 20$. The only difference between the two models is the forecast window in the response variable. However, we immediately see that the significance of the 3rd and 4th lag of the VOI factor and the 2nd lag of the OIR factor is more prominent – the coefficient is negative and significant for a larger proportion of the year. Lastly, using model (3.8), the correlation between the actual and predicted price change is 0.434 and using the predicted price change as a trinomial class variable, the correlation is 0.758, giving very large improvements from the correlations in model (2.3).

The results of the trading strategy with using the linear model (3.8) and trading parameters $q = 0.15, w = (0.5, 0.5)$ is summarized in Table 3.5 below. The full results of this trading strategy can be found in appendix A.3.

Statistical test: one-tailed, one-sample t-test (df = 243) $H_0 : \mu \geq 0$ $H_1 : \mu < 0$						
Mean daily profit (CNY)	Standard error	t-stat	p-value	Days with profit	Days with losses	Average daily trade volume
58,600	8,091	7.2431	2.886×10^{-12}	231	12	1,798
Average Daily Sharpe Ratio: 0.464 Annualized Sharpe Ratio: 7.243						

Table 3.5: Trading strategy results using improved linear model

For every parameter pair (k, q), we perform the following paired one-tailed t-test to find the confidence perimeter:

$$\begin{aligned} H_0 &: \mu_{5,0.15} - \mu_{k,q} \leq 0 \\ H_1 &: \mu_{5,0.15} - \mu_{k,q} > 0 \end{aligned} \tag{3.9}$$

By calculating the t-statistic and p-value for every mesh point, we can find a perimeter in which the result becomes statistically insignificant at the 90%, 95%, and 99% confidence levels. We would thereby obtain the (k, q) pairs that form this confidence perimeter and conclude that the true maximum mean daily profit occurs when (k, q) lies entirely within the perimeter. The heatmap for the 90%, 95%, and 99% confidence regions are shown in Figure 3.9 below. The red cell is the benchmark at which we apply the above paired t-test (3.9) to the daily profits. The darkest blue cells represent the parameter pair at which we cannot reject the null hypothesis at the 99% confidence level. Similarly, the medium blue cells represent the 95% confidence level, and the lightest blue cells represent the 90% confidence level.

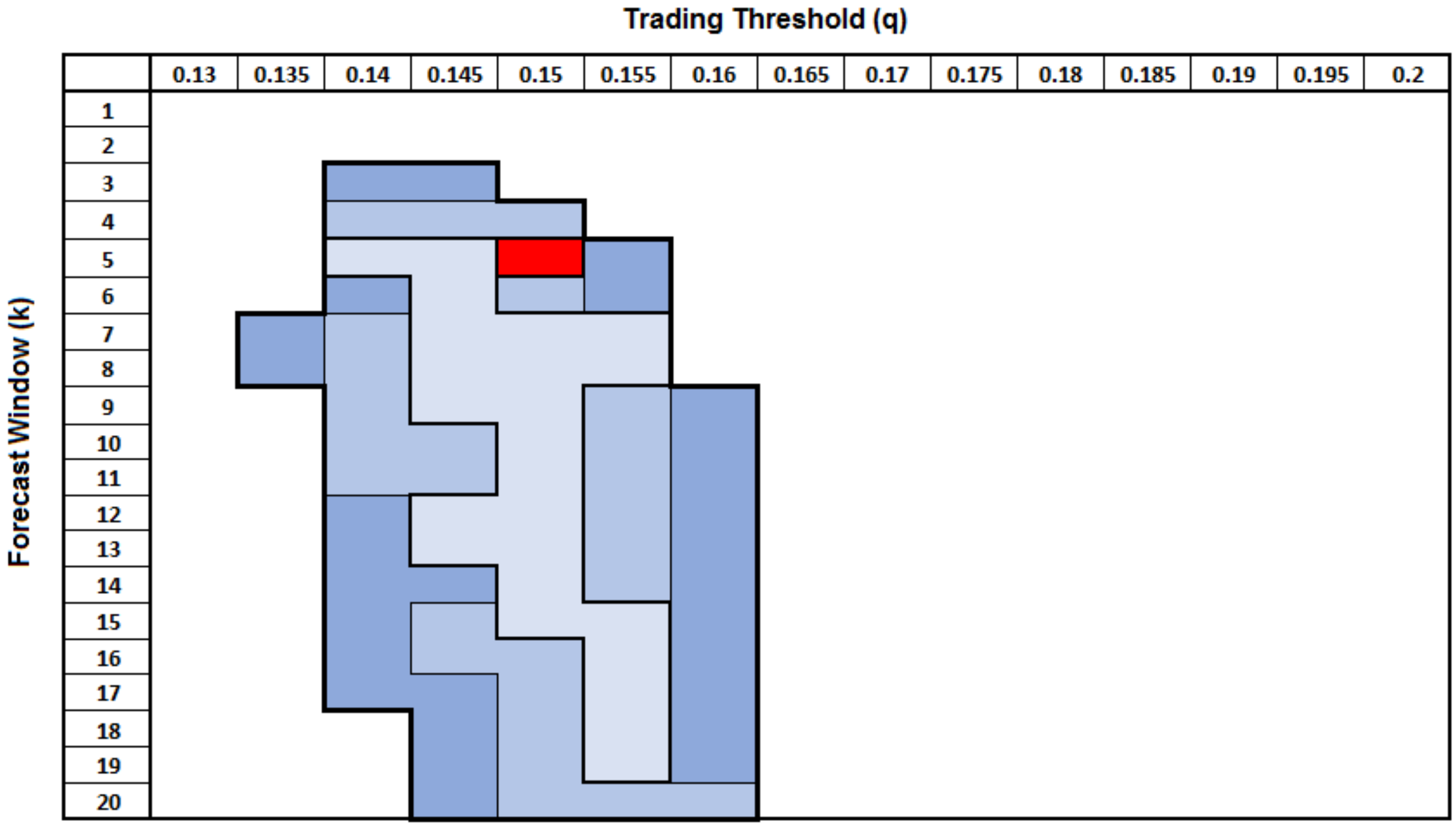

Figure 3.9: Confidence region heatmap for the strategy using the linear model (3.6) with coefficient weights $w_1 = w_2 = 0.5$ over the mesh $q = 0.13, 0.135, ..., 0.2$ and $k = 1, ..., 20$. Red cell indicates optimal strategy. The 90%, 95%, and 99% confidence regions are represented by light blue, medium blue, and dark blue cells respectively. Note that even at the 99% confidence region, the optimal threshold q lies between 0.135 and 0.16. The forecast window, however, has a large interval – from 3 to 20.

Based on the heatmap above, we find that the maximum daily profit can occur:

- At the 90% confidence level, for parameters $5 \leq k \leq 19$ and $0.14 \leq q \leq 0.155$
- At the 95% confidence level, for parameters $4 \leq k \leq 20$ and $0.14 \leq q \leq 0.16$
- At the 99% confidence level, for parameters $3 \leq k \leq 20$ and $0.135 \leq q \leq 0.16$.

Although this optimization method is not ideal since changing the lag L on the VOI and OIR variables or changing the coefficient weights w_j can impact the selection of the forecast window and trading threshold. But based on the some preliminary test results (see appendix B for details), for lags $L = 2$, $L = 3$, and $L = 4$ the optimal parameters (k, q) are $(5, 0.15)$, $(6, 0.15)$, and $(9, 0.15)$ respectively. As these parameters lie in our 95% confidence region for the lag-5 results in Figure 3.9 above, we cannot reject the null hypothesis and can only concude that lag-selection does not statistically impact the choice of k and q.

As our strategy is highly successful and each trade execution generates positive profit on average, these parameter results make sense. We cannot expect to earn a greater profit by increasing the trading threshold as it would only decrease the amount of trades we make. Although economically speaking, decreasing the threshold to less than 0.2 means that we could potentially sell below buy price, it seems our signal is strong enough such that even if we forecast a 0.15 mid-price change, the actual mid-price change is often greater. We find that for $k = 5$, on average 60% of mid-price changes greater than 0.15 are also greater than 0.20. By decreasing the threshold to 0.15, the strategy trades more often and thus generates a larger profit.

3.3 Summary and Final Considerations

The additional factors OIR and MPB and the spread adjustment to all factors have improved the daily profit by over 350% to 50,369 CNY and the win ratio by nearly 17% compared to the original strategy using the volume order imbalance model (2.3). The model and strategy is then further improved by selecting the optimal regression and trading parameters. First, we found that due to the strong positive autocorrelation of the daily coefficients of the linear model, using a 2-day simple moving average had a statistically significant increase in profit. We also found that the lag-5 model on VOI and OIR was the simplest model that significantly outperformed others. Lastly, for the remaining parameters, we showed a confidence region in which the optimal forecast window and trading threshold could lie at the 90%, 95% and 99% level. The optimal parameters were found to be around $k = 5$ and $q = 0.15$ giving an average daily profit of 58,600 CNY. Finally, we saw a large improvement in the correlation between the actual price change and the predicted price change confirming that this strategy makes the correct trade more often.

There are still many important considerations that must be taken into account to verify the validity of this strategy. Of the assumptions stated in section 2.2, (a) and (b) are not completely realistic. The financial markets are not devoid of competing traders so the assumption of always being able to take the best counterparty price (sell at bid, buy at ask) is not valid. It is unlikely that we will always be able to take the best counterparty price for every trading signal we receive. Although computers are able to receive market data, compute the trading signal using the model coefficients, and make a trading decision within milliseconds, by the time the order is sent back to the exchange, the actual execution price might already have changed. Even executing a few milliseconds earlier can result in more profitable trades [14]. We could improve the accuracy of this trading simulation by assuming we are able to take the counterparty price 50% of the time or building a program to model our competitors in the markets.

Finally, we ask whether this strategy can be tricked by our competitors. Given that order imbalance is not an unknown technique as a trading strategy, there will be firms who will try to take advantage of traders using this trading signal. By looking carefully at the three factors we use to generate our trading signal: volume order imbalance (VOI), order imbalance ratio (OIR), and mid-price basis (MPB), we find that VOI and OIR heavily rely on the volume of the best bid and best ask prices. Competitors can manipulate these figures by quickly submitting a large buy or sell limit order and cancelling immediately after. If our trading algorithm picks up a large bid or ask volume due to the competitor's spoofing technique, we would incorrectly calculate a large order imbalance (both VOI and OIR) and end up trading on a signal that is falsely generated. We will not be covering the topic of making the trading algorithm more robust to handle competitor spoofing as it is not within the scope of this paper.

Chapter 4

Conclusions

We first introduced the area of high frequency trading and the data that will be used to test the trading strategy. By examining order imbalance, a measure of the difference in size of buy and sell orders in the market, we developed a simple trading strategy by fitting a linear model using ordinary least squares against a 20 time-step (10 second) average price change. We have shown that the strategy, using this linear model to forecast future price changes and trading when the forecast is greater than 0.2 ticks, is highly profitable. However, after some analysis was done on the profits made by the strategy, we found much of it was strongly positively correlated with the total trading volume in the market. We further improved on this trading signal by extending the linear model to include 2 new factors: order imbalance ratio, a measure of the degree of imbalance, and the mid-price basis, a mean-reverting process. All three factors were also adjusted by dividing by the bid-ask spread as we found that on most days, large spreads indicated low price changes. Lastly, we determined a confidence interval for the optimal regression and trading parameters: the forecast window for the average price change and the trading threshold and found that they were closer to 5 and 0.15 respectively.

4.1 Further Work

There are several areas of research that may improve the robustness and profitability of this trading strategy. We outline two ideas below.

The study by Cont [5] analyzed the relationship between order flow imbalance and intraday volatility (diurnal). They found that the market depth during the 30 minutes after market open is quite shallow meaning that orders submitted by traders during that time can have a large impact on price movement. However, Huang [7] also attempted find a relationship between order imbalance and volatility by using

GARCH(1,1) but they found no such relationship. As these results appear to be inconsistent, this area can be explored further and either be used to enhance our existing signal or used to create another trading signal.

Further model enhancements could also be done. In addition to linear regressors, we could model the response variable (k-step average price change) with a time-series AR($k+1$) model. Preliminary results indicate that the both the autocorrelation and partial autocorrelation in the response is significant past $k+1$ lags meaning we could potentially use the lag-($k+1$) term as a feature in our model. The reason we cannot choose lags $1, ..., k$ is because those are not known by the time the market data is received since the response depends on data k-steps ahead. Lastly, we may want to use more sophisticated statistical techniques to do model selection, including machine learning or lasso regression while bearing in mind that a better statistical model does not necessarily translate to better profits.

The final suggestion would be to apply machine learning techniques to make trading decisions. As mentioned in the previous chapters, the predicted price change was essentially a trinomial classifier for the trading strategy based on the trading threshold q. We can take advantage of the high correlation between the trinomial variable and the actual price change. Instead of using the linear model to forecast a continuous variable, we could apply one of several machine learning techniques to build a trinomial classifier, such as logistic regression, support vector machines, or random forests. However, given the vast amount of data in high frequency trading, it would be important to split the data into training and testing sets to ensure the trading strategy does not make decisions using an overfitted classifier.

Bibliography

[1] Irene Aldridge. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. John Wiley & Sons, Inc, 2010.

[2] Amélie Charles and Olivier Darné. Variance ratio tests of random walk: An overview. *Journal of Economic Surveys*, 23(3):503–527, 2009.

[3] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, 65:111–130, 2002.

[4] Tarun Chordia and Avanidhar Subrahmanyam. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 72:485–518, 2004.

[5] Rama Cont, Arseniy Kukanov, and Sasha Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2014.

[6] Martin D. Gould, Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

[7] Han-Ching Huang, Yong-Chern Su, and Yi-Chun Liu. The performance of imbalance-based trading strategy on tender offer announcement day. *Investment Management and Financial Innovations*, 11(2):38–46, 2014.

[8] Jonathan Karpoff. The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1):109–126, 1987.

[9] Charles M.C. Lee and Mark J. Ready. Inferring trade direction from intraday data. *The Journal of Finance*, 46:733–747, 1991.

[10] Xuan Liu. Stationary processes and hft. Master's thesis, University of Oxford, 2015.

[11] Rahul Ravi and Yuqing Sha. Autocorrelated Order-Imbalance and Price Momentum in the Stock Market. *International Journal of Economics and Finance*, 6(10):39–54, 2014.

[12] Ruey S. Tsay. *Analysis of Financial Time Series.* John Wiley & Sons, Inc, 2010.

[13] Geroge H.K. Wang and Jot Yau. Trading volume, bid-ask spread, and price volatility in futures markets. *Journal of Futures Markets*, 20(10):943–970, 2000.

[14] Zhaodong Wang and Weian Zheng. *High-Frequency Trading and Probability Theory.* World Scientific Publishing Co. Pte. Ltd, 2015.

Appendix A

Daily Strategy P&L

A.1 Volume Order Imbalance Strategy P&L: Main Contract

Total days with profit: 185
Total days with losses: 58
Average Daily Sharpe Ratio: 0.380
Annualized Sharpe Ratio: 5.935
* all numbers reported in CNY

Average:	10,078	9,449	19,528	634	68.78
Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-01-02	0	0	0	0	0
2014-01-03	8057	2326	10383	198	17.09
2014-01-06	-90	-161	-251	296	24.96
2014-01-07	9535	-291	9245	500	41.98
2014-01-08	12836	596	13432	310	26.14
2014-01-09	6700	4697	11396	456	38.42
2014-01-10	13036	6157	19193	502	41.74
2014-01-13	-2700	9286	6586	414	34.27
2014-01-14	1106	-4556	-3450	372	30.75
2014-01-15	-1164	27	-1137	94	7.78
2014-01-16	7923	140	8063	96	7.98
2014-01-17	-3760	-659	-4419	206	17
2014-01-20	-6270	763	-5507	332	27.23
2014-01-21	-3411	-2593	-6004	298	24.62
2014-01-22	10460	-2965	7495	240	20.12
2014-01-23	-4730	198	-4532	180	15.16
2014-01-24	-2211	2062	-149	62	5.24
2014-01-27	-2004	-1011	-3015	166	13.88
2014-01-28	329	557	887	198	16.57
2014-01-29	-2910	-2231	-5141	124	10.41
2014-01-30	3294	7675	10969	98	8.16
2014-02-07	-4544	7677	3134	174	14.33
2014-02-10	13129	3071	16200	220	18.6
2014-02-11	12940	-4032	8908	206	17.56
2014-02-12	1965	-1259	706	92	7.87

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-02-13	-4668	-6092	-10760	152	13
2014-02-14	-1139	-582	-1722	34	2.91
2014-02-17	5193	-789	4403	16	1.38
2014-02-18	-1074	-890	-1964	90	7.72
2014-02-19	943	612	1555	244	21.02
2014-02-20	12420	-2303	10117	356	30.72
2014-02-21	-1761	-2085	-3846	328	27.93
2014-02-24	4433	6251	10684	94	7.78
2014-02-25	2765	21149	23914	454	37.33
2014-02-26	8192	-6909	1284	550	44.28
2014-02-27	5483	7887	13371	606	48.95
2014-02-28	-277	10546	10270	520	41.95
2014-03-03	11857	-4360	7496	546	44.42
2014-03-04	5003	5718	10721	406	32.79
2014-03-05	7905	6406	14310	428	34.65
2014-03-06	10695	6876	17571	516	41.45
2014-03-07	14930	-982	13948	672	54.46
2014-03-10	4806	3913	8720	422	33.2
2014-03-11	14284	-3952	10332	888	69.24
2014-03-12	11208	7963	19171	574	44.85
2014-03-13	14250	10062	24312	736	58.14
2014-03-14	4275	-1355	2920	482	37.9
2014-03-17	7915	-1087	6828	148	11.76
2014-03-18	1168	-5225	-4057	924	73.68
2014-03-19	-309	-1105	-1414	86	6.77
2014-03-20	13660	8437	22097	358	28.12
2014-03-21	-1758	43880	42122	662	52.19
2014-03-24	9340	9276	18617	1202	96.82
2014-03-25	12579	3554	16133	602	48.63
2014-03-26	3525	2412	5937	312	25.22
2014-03-27	-7230	14603	7373	324	26.13
2014-03-28	10408	6301	16709	526	42.45
2014-03-31	759	8403	9162	574	46.29
2014-04-01	5807	4816	10623	334	26.98
2014-04-02	1348	4234	5582	310	25.19
2014-04-03	-1798	1469	-329	156	12.74
2014-04-04	4123	-1440	2683	248	20.18
2014-04-08	12831	14901	27732	244	20.34
2014-04-09	-3002	-335	-3337	376	31.69
2014-04-10	-2806	6943	4137	170	14.41
2014-04-11	1392	521	1912	324	27.64
2014-04-14	2719	2738	5456	180	15.32
2014-04-15	5317	-7489	-2172	154	12.96
2014-04-16	9684	4378	14062	324	27.19
2014-04-17	5449	274	5723	124	10.39
2014-04-18	12421	2215	14636	224	18.62
2014-04-21	19988	5626	25614	780	64.83
2014-04-22	4348	10778	15126	502	41.07
2014-04-23	10467	3315	13782	250	20.49
2014-04-24	1030	-1490	-461	222	18.2
2014-04-25	-3741	1524	-2216	172	14.08

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-04-28	-3206	2103	-1103	366	29.52
2014-04-29	-2364	525	-1839	352	28.39
2014-04-30	-1196	1998	802	32	2.59
2014-05-05	5512	5354	10866	22	1.77
2014-05-06	-307	-475	-782	334	27.01
2014-05-07	123	2104	2227	216	17.36
2014-05-08	9487	2975	12462	262	21.09
2014-05-09	446	907	1353	264	21.13
2014-05-12	-55	3738	3684	338	27.48
2014-05-13	-6729	-3831	-10560	372	30.3
2014-05-14	-4538	1610	-2928	14	1.14
2014-05-15	-267	-2230	-2496	62	4.98
2014-05-16	-6534	-3717	-10250	108	8.65
2014-05-19	5700	3567	9267	120	9.47
2014-05-20	3862	-6583	-2721	324	25.61
2014-05-21	4097	311	4408	222	17.56
2014-05-22	20447	-1625	18823	290	23.19
2014-05-23	-1810	3948	2138	306	24.41
2014-05-26	5337	-2935	2402	134	10.79
2014-05-27	3460	3572	7033	152	12.24
2014-05-28	1507	1863	3371	152	12.25
2014-05-29	-1943	-234	-2177	212	17.18
2014-05-30	1662	-5558	-3896	48	3.88
2014-06-03	1906	1598	3504	58	4.68
2014-06-04	3038	-403	2635	102	8.12
2014-06-05	-2346	1932	-414	116	9.27
2014-06-06	3631	6293	9924	100	7.98
2014-06-09	-7626	3797	-3829	198	15.85
2014-06-10	7686	8758	16444	410	32.98
2014-06-11	-2102	-3723	-5825	428	34.52
2014-06-12	84	-696	-612	64	5.16
2014-06-13	3450	2237	5687	60	4.87
2014-06-16	-385	2090	1705	196	16.08
2014-06-17	-6599	-3032	-9630	68	5.55
2014-06-18	-1418	525	-893	18	1.46
2014-06-19	-2996	2785	-211	88	7.05
2014-06-20	124	4635	4759	160	12.81
2014-06-24	-1522	-11	-1533	126	10.06
2014-06-25	3565	1123	4689	120	9.56
2014-06-26	1290	3184	4474	226	18.13
2014-06-27	-584	4278	3695	110	8.83
2014-06-30	6719	-554	6164	90	7.28
2014-07-01	6648	587	7236	150	12.12
2014-07-02	1813	3327	5140	120	9.7
2014-07-03	2045	-1769	276	90	7.32
2014-07-04	-305	-875	-1179	80	6.5
2014-07-07	4060	1480	5540	32	2.6
2014-07-08	816	1065	1881	32	2.59
2014-07-09	-76	4567	4491	86	6.94
2014-07-10	-164	-416	-581	40	3.2
2014-07-11	649	1777	2426	88	7.07

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-07-14	-881	114	-766	40	3.23
2014-07-15	-3649	-2864	-6513	76	6.17
2014-07-16	-3385	1103	-2283	26	2.11
2014-07-17	-2806	1016	-1791	18	1.45
2014-07-18	-1570	-2160	-3730	256	20.75
2014-07-21	-366	-1570	-1936	64	5.18
2014-07-22	4317	3043	7360	60	4.9
2014-07-23	1855	453	2308	188	15.46
2014-07-24	5924	-51	5873	130	10.83
2014-07-25	1500	6324	7824	340	28.68
2014-07-28	16003	1120	17123	288	25.08
2014-07-29	6392	7479	13871	946	83.04
2014-07-30	12254	-6086	6168	374	32.76
2014-07-31	6306	-828	5479	344	30.21
2014-08-01	5976	19497	25473	262	23.24
2014-08-04	10986	6686	17672	654	58.04
2014-08-05	6223	8460	14683	648	57.68
2014-08-06	1037	4490	5527	556	49.16
2014-08-07	12195	12035	24230	364	32.15
2014-08-08	3775	2180	5955	444	38.93
2014-08-11	9443	204	9647	258	22.86
2014-08-12	538	750	1288	420	37.06
2014-08-13	11373	-2036	9337	178	15.71
2014-08-14	4091	7657	11748	380	33.66
2014-08-15	5323	2288	7612	406	36.04
2014-08-18	14154	585	14739	526	46.91
2014-08-19	8861	1455	10316	398	35.42
2014-08-20	-2378	3835	1457	222	19.72
2014-08-21	2600	2102	4702	254	22.39
2014-08-22	246	5054	5299	192	17
2014-08-25	2405	4093	6498	264	23.31
2014-08-26	7947	77	8024	216	18.98
2014-08-27	1435	-908	528	134	11.76
2014-08-28	-1152	1692	540	120	10.5
2014-08-29	-630	-844	-1474	108	9.47
2014-09-01	4491	-2154	2337	92	8.11
2014-09-02	-1891	6617	4726	146	12.97
2014-09-03	1507	366	1874	208	18.83
2014-09-04	-4454	8452	3998	318	28.91
2014-09-05	1268	-1153	115	212	19.52
2014-09-09	4737	7125	11862	232	21.39
2014-09-10	-2743	340	-2402	174	15.91
2014-09-11	8620	9644	18264	146	13.38
2014-09-12	1424	6222	7646	568	51.77
2014-09-15	1476	158	1634	134	12.23
2014-09-16	7676	5616	13292	196	17.84
2014-09-17	-4537	950	-3587	176	15.84
2014-09-18	10991	139	11129	360	32.55
2014-09-19	10407	1454	11861	268	24.4
2014-09-22	4820	7272	12092	548	49.34
2014-09-23	3547	8860	12407	500	45.07

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-09-24	8679	2806	11484	360	32.88
2014-09-25	2781	16183	18963	572	53.08
2014-09-26	5739	4892	10631	636	58.44
2014-09-29	3378	5742	9120	486	44.7
2014-09-30	-169	-1790	-1959	230	21.2
2014-10-08	-6806	5460	-1346	196	18.13
2014-10-09	5121	804	5925	278	25.88
2014-10-10	536	2152	2688	250	23.16
2014-10-13	3435	614	4048	146	13.36
2014-10-14	1881	3175	5056	372	34.12
2014-10-15	4390	6941	11331	128	11.75
2014-10-16	1164	1668	2833	256	23.64
2014-10-17	4797	11118	15915	516	47.32
2014-10-20	1944	4025	5969	194	17.85
2014-10-21	7242	1510	8751	180	16.54
2014-10-22	3524	-1353	2171	196	17.95
2014-10-23	2267	3634	5901	98	8.9
2014-10-24	1857	2782	4639	112	10.11
2014-10-27	3165	-3080	86	86	7.67
2014-10-28	-3318	1354	-1964	252	22.72
2014-10-29	2521	-1732	789	114	10.45
2014-10-30	2282	4636	6918	220	20.31
2014-10-31	4961	15194	20154	292	27.33
2014-11-03	2905	2466	5372	430	40.64
2014-11-04	907	1157	2064	158	14.88
2014-11-05	-4212	-3837	-8049	112	10.54
2014-11-06	111	2211	2322	100	9.39
2014-11-07	3541	10257	13798	254	24.01
2014-11-10	9412	4139	13551	730	69.35
2014-11-11	17337	17112	34450	698	67.15
2014-11-12	-560	3614	3054	866	83.13
2014-11-13	657	-3074	-2417	408	39.38
2014-11-14	671	9517	10188	428	41.16
2014-11-17	5929	8176	14106	322	31.17
2014-11-18	3800	3194	6994	726	69.43
2014-11-19	-795	-725	-1519	86	8.2
2014-11-20	-1726	-4571	-6297	280	26.68
2014-11-21	8343	9881	18224	228	21.98
2014-11-24	15880	9317	25197	608	60.61
2014-11-25	-1434	12914	11480	1102	111.2
2014-11-26	18841	13285	32126	590	60.47
2014-11-27	8690	17374	26064	914	94.08
2014-11-28	3060	34733	37793	686	71.13
2014-12-01	20810	23580	44390	1214	128.45
2014-12-02	13856	73994	87850	1280	137.65
2014-12-03	168577	82763	251340	2196	247.2
2014-12-04	116192	128421	244613	4540	524.73
2014-12-05	217004	113837	330841	3340	398.69
2014-12-08	108507	101458	209965	6230	755.27
2014-12-09	132763	222644	355407	5554	693.47
2014-12-10	239180	104059	343239	6188	748.91

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-12-11	55994	81541	137535	7358	890.62
2014-12-12	82522	27095	109618	5136	623.11
2014-12-15	5708	-15445	-9737	3418	411.68
2014-12-16	77784	28768	106552	2414	296.44
2014-12-17	67746	21052	88798	1842	230.11
2014-12-18	25066	107545	132610	3274	428.35
2014-12-19	57666	77215	134881	4444	577.8
2014-12-22	42810	95357	138167	3744	488.17
2014-12-23	19916	49924	69840	4726	604.8
2014-12-24	52519	54053	106572	3626	444.84
2014-12-25	12081	67774	79855	4084	504.92
2014-12-26	53049	94586	147635	3532	452.82
2014-12-29	71137	36480	107616	4022	532.92
2014-12-30	82006	48640	130646	4372	571.67
2014-12-31	2867	23121	25987	3998	528.86

Table A.1: Daily P&L results for volume order imbalance strategy trading main futures contract. Pink cells indicate a loss during the session or entire day.

A.2 Volume Order Imbalance Strategy P&L: Secondary Contract

Total days with profit: 119 Total days with losses: 99 Average Daily Sharpe Ratio: 0.113 Annualized Sharpe Ratio: 1.763 * all numbers reported in CNY					
Average:	1,339	1,462	2,801	255	28.22
Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-01-02	0	0	0	0	0
2014-01-03	7225	325	7551	4	0.35
2014-01-06	14546	-574	13972	4	0.34
2014-01-07	-9824	473	-9351	28	2.36
2014-01-08	2325	2453	4778	12	1.01
2014-01-09	5932	5306	11239	6	0.51
2014-01-10	0	0	0	0	0
2014-01-13	227	2508	2735	34	2.82
2014-01-14	-3400	113	-3286	10	0.83
2014-01-15	-2426	2301	-126	22	1.83
2014-01-16	-33	1580	1546	8	0.67
2014-01-17	9621	-551	9070	14	1.15
2014-01-20	-5750	-2242	-7992	286	23.46
2014-01-21	-3411	-2593	-6004	298	24.62
2014-01-22	10460	-2965	7495	240	20.12
2014-01-23	-4730	198	-4532	180	15.16
2014-01-24	-2211	2062	-149	62	5.24

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-01-27	-2004	-1011	-3015	166	13.88
2014-01-28	329	557	887	198	16.57
2014-01-29	-2910	-2231	-5141	124	10.41
2014-01-30	3294	7675	10969	98	8.16
2014-02-07	-1953	-4093	-6046	10	0.83
2014-02-10	0	0	0	0	0
2014-02-11	0	0	0	0	0
2014-02-12	-1029	0	-1029	4	0.34
2014-02-13	0	1166	1166	2	0.17
2014-02-14	26	0	26	2	0.17
2014-02-17	0	0	0	0	0
2014-02-18	1217	1714	2931	18	1.55
2014-02-19	-5211	2106	-3105	34	2.93
2014-02-20	-902	3326	2424	16	1.38
2014-02-21	-3694	1529	-2166	18	1.53
2014-02-24	-4400	-2015	-6415	152	12.57
2014-02-25	2765	21149	23914	454	37.33
2014-02-26	8192	-6909	1284	550	44.28
2014-02-27	5483	7887	13371	606	48.95
2014-02-28	-277	10546	10270	520	41.95
2014-03-03	687	0	687	2	0.16
2014-03-04	-5857	3983	-1874	12	0.97
2014-03-05	-351	-125	-476	22	1.78
2014-03-06	0	-1357	-1357	6	0.49
2014-03-07	-1295	-3171	-4465	46	3.73
2014-03-10	-4292	0	-4292	2	0.16
2014-03-11	-9302	0	-9302	4	0.31
2014-03-12	441	-5120	-4679	100	7.79
2014-03-13	-9328	-4449	-13777	48	3.78
2014-03-14	-1871	706	-1165	40	3.12
2014-03-17	0	0	0	0	0
2014-03-18	854	-5392	-4538	44	3.49
2014-03-19	1109	5369	6477	4	0.31
2014-03-20	-11827	4949	-6878	10	0.79
2014-03-21	-1417	10695	9279	24	1.91
2014-03-24	3592	-918	2675	266	21.43
2014-03-25	12579	3554	16133	602	48.63
2014-03-26	3525	2412	5937	312	25.22
2014-03-27	-7230	14603	7373	324	26.13
2014-03-28	10408	6301	16709	526	42.45
2014-03-31	759	8403	9162	574	46.29
2014-04-01	0	0	0	0	0
2014-04-02	-3532	-3882	-7414	28	2.27
2014-04-03	-873	0	-873	2	0.16
2014-04-04	-6365	-1385	-7750	8	0.65
2014-04-08	1335	-3155	-1820	30	2.5
2014-04-09	-3221	0	-3221	6	0.5
2014-04-10	-1414	-6454	-7868	4	0.34
2014-04-11	-5014	-1414	-6428	4	0.34
2014-04-14	-4722	1046	-3676	8	0.68
2014-04-15	0	0	0	0	0

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-04-16	9965	-3989	5976	30	2.52
2014-04-17	1820	3712	5532	96	8.04
2014-04-18	-26	-1193	-1219	30	2.49
2014-04-21	4839	2186	7026	90	7.47
2014-04-22	4348	10778	15126	502	41.07
2014-04-23	10467	3315	13782	250	20.49
2014-04-24	1030	-1490	-461	222	18.2
2014-04-25	-3741	1524	-2216	172	14.08
2014-04-28	-3206	2103	-1103	366	29.52
2014-04-29	-2364	525	-1839	352	28.39
2014-04-30	-1196	1998	802	32	2.59
2014-05-05	0	0	0	0	0
2014-05-06	0	0	0	0	0
2014-05-07	-5808	-2268	-8076	36	2.88
2014-05-08	-9309	5188	-4121	10	0.8
2014-05-09	-692	0	-692	2	0.16
2014-05-12	6443	862	7305	12	0.97
2014-05-13	422	1144	1566	44	3.57
2014-05-14	0	-3905	-3905	4	0.32
2014-05-15	1495	203	1698	10	0.81
2014-05-16	0	868	868	2	0.16
2014-05-19	-1829	-324	-2153	68	5.37
2014-05-20	3862	-6583	-2721	324	25.61
2014-05-21	4097	311	4408	222	17.56
2014-05-22	20447	-1625	18823	290	23.19
2014-05-23	-1810	3948	2138	306	24.41
2014-05-26	5337	-2935	2402	134	10.79
2014-05-27	3460	3572	7033	152	12.24
2014-05-28	1507	1863	3371	152	12.25
2014-05-29	-1943	-234	-2177	212	17.18
2014-05-30	1662	-5558	-3896	48	3.88
2014-06-03	0	0	0	0	0
2014-06-04	2068	-1892	177	4	0.32
2014-06-05	0	0	0	0	0
2014-06-06	0	2456	2456	4	0.32
2014-06-09	1768	-572	1196	4	0.32
2014-06-10	5144	-3187	1957	50	4.01
2014-06-11	-4657	628	-4029	8	0.64
2014-06-12	0	0	0	0	0
2014-06-13	6268	0	6268	2	0.16
2014-06-16	747	-4506	-3758	6	0.49
2014-06-17	87	0	87	2	0.16
2014-06-18	0	0	0	0	0
2014-06-19	-725	0	-725	4	0.32
2014-06-20	1648	-1491	157	54	4.31
2014-06-24	-1703	2712	1009	42	3.36
2014-06-25	3565	1123	4689	120	9.56
2014-06-26	1290	3184	4474	226	18.13
2014-06-27	-584	4278	3695	110	8.83
2014-06-30	6719	-554	6164	90	7.28
2014-07-01	0	0	0	0	0

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-07-02	0	0	0	0	0
2014-07-03	0	-1293	-1293	2	0.16
2014-07-04	0	0	0	0	0
2014-07-07	0	0	0	0	0
2014-07-08	0	0	0	0	0
2014-07-09	0	0	0	0	0
2014-07-10	-692	-692	-1384	4	0.32
2014-07-11	0	0	0	0	0
2014-07-14	-332	-605	-937	6	0.49
2014-07-15	0	-1597	-1597	6	0.49
2014-07-16	0	0	0	0	0
2014-07-17	-402	-2792	-3194	12	0.97
2014-07-18	-3937	-1208	-5145	36	2.92
2014-07-21	-2119	-37	-2156	22	1.78
2014-07-22	4317	3043	7360	60	4.9
2014-07-23	1855	453	2308	188	15.46
2014-07-24	5924	-51	5873	130	10.83
2014-07-25	1500	6324	7824	340	28.68
2014-07-28	16003	1120	17123	288	25.08
2014-07-29	6392	7479	13871	946	83.04
2014-07-30	12254	-6086	6168	374	32.76
2014-07-31	6306	-828	5479	344	30.21
2014-08-01	-756	-13776	-14531	4	0.36
2014-08-04	-1624	-1187	-2810	50	4.45
2014-08-05	-2993	-3282	-6275	86	7.68
2014-08-06	-834	3123	2290	40	3.55
2014-08-07	-5073	1998	-3076	82	7.28
2014-08-08	-4669	-2547	-7217	76	6.68
2014-08-11	-897	2377	1480	18	1.6
2014-08-12	-6809	-2610	-9419	44	3.89
2014-08-13	6692	-6626	66	64	5.67
2014-08-14	5555	-1221	4334	40	3.53
2014-08-15	-326	-3411	-3737	52	4.59
2014-08-18	4973	-2588	2385	650	57.97
2014-08-19	8861	1455	10316	398	35.42
2014-08-20	-2378	3835	1457	222	19.72
2014-08-21	2600	2102	4702	254	22.39
2014-08-22	246	5054	5299	192	17
2014-08-25	2405	4093	6498	264	23.31
2014-08-26	7947	77	8024	216	18.98
2014-08-27	1435	-908	528	134	11.76
2014-08-28	-1152	1692	540	120	10.5
2014-08-29	-630	-844	-1474	108	9.47
2014-09-01	0	0	0	0	0
2014-09-02	-731	-6791	-7522	18	1.61
2014-09-03	8362	1344	9705	14	1.27
2014-09-04	-1358	5031	3673	42	3.83
2014-09-05	-4851	3867	-984	24	2.22
2014-09-09	715	-6167	-5452	32	2.96
2014-09-10	3263	0	3263	2	0.18
2014-09-11	6664	5830	12494	22	2.03

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-09-12	-1042	-2523	-3565	90	8.23
2014-09-15	8577	-2473	6104	14	1.28
2014-09-16	-4130	935	-3195	14	1.28
2014-09-17	-2725	-721	-3446	48	4.33
2014-09-18	3204	1464	4668	4	0.36
2014-09-19	-2413	-1263	-3676	34	3.08
2014-09-22	-2766	4285	1519	332	29.91
2014-09-23	3547	8860	12407	500	45.07
2014-09-24	8679	2806	11484	360	32.88
2014-09-25	2781	16183	18963	572	53.08
2014-09-26	5739	4892	10631	636	58.44
2014-09-29	3378	5742	9120	486	44.7
2014-09-30	-169	-1790	-1959	230	21.2
2014-10-08	0	-2797	-2797	2	0.19
2014-10-09	2745	1103	3848	6	0.56
2014-10-10	0	0	0	0	0
2014-10-13	4043	-1190	2853	8	0.74
2014-10-14	1856	-5343	-3488	20	1.84
2014-10-15	7583	2372	9956	10	0.92
2014-10-16	-3111	0	-3111	6	0.55
2014-10-17	1544	-2262	-718	36	3.29
2014-10-20	-1696	-1241	-2937	378	34.79
2014-10-21	7242	1510	8751	180	16.54
2014-10-22	3524	-1353	2171	196	17.95
2014-10-23	2267	3634	5901	98	8.9
2014-10-24	1857	2782	4639	112	10.11
2014-10-27	3165	-3080	86	86	7.67
2014-10-28	-3318	1354	-1964	252	22.72
2014-10-29	2521	-1732	789	114	10.45
2014-10-30	2282	4636	6918	220	20.31
2014-10-31	4961	15194	20154	292	27.33
2014-11-03	-2754	0	-2754	6	0.57
2014-11-04	-998	-653	-1651	8	0.75
2014-11-05	-2618	-938	-3555	4	0.38
2014-11-06	-1553	-7215	-8768	10	0.94
2014-11-07	2555	4652	7207	44	4.17
2014-11-10	-294	4136	3842	110	10.49
2014-11-11	5980	-3885	2095	134	12.93
2014-11-12	-1800	-3376	-5176	116	11.18
2014-11-13	-4758	-1535	-6292	12	1.16
2014-11-14	1157	2610	3767	38	3.66
2014-11-17	12965	-6487	6478	130	12.61
2014-11-18	1093	-2058	-964	116	11.12
2014-11-19	501	2050	2551	102	9.74
2014-11-20	0	0	0	0	0
2014-11-21	4522	-175	4346	8	0.77
2014-11-24	0	0	0	0	0
2014-11-25	-6480	-15298	-21778	270	27.29
2014-11-26	478	8277	8755	44	4.52
2014-11-27	-11437	-3344	-14781	376	38.81
2014-11-28	0	-12222	-12222	2	0.21

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-12-01	-4200	-962	-5162	156	16.51
2014-12-02	10641	21361	32003	46	4.99
2014-12-03	-23221	-20922	-44144	752	84.82
2014-12-04	-25761	-30395	-56156	776	90.28
2014-12-05	-11478	1041	-10437	230	27.59
2014-12-08	-49925	-41967	-91892	774	94.36
2014-12-09	20690	-92473	-71783	116	14.51
2014-12-10	-94647	-37320	-131967	1438	175.33
2014-12-11	-32035	-57564	-89599	468	57.39
2014-12-12	16686	-3412	13274	124	15.23
2014-12-15	-24622	11151	-13471	178	21.76
2014-12-16	-21529	4123	-17406	58	7.23
2014-12-17	63559	46585	110144	1054	135.38
2014-12-18	-7668	1913	-5755	432	54.58
2014-12-19	14855	12668	27523	70	8.78
2014-12-22	27659	69559	97218	2162	281.91
2014-12-23	19916	49924	69840	4726	604.8
2014-12-24	52519	54053	106572	3626	444.84
2014-12-25	12081	67774	79855	4084	504.92
2014-12-26	53049	94586	147635	3532	452.82
2014-12-29	71137	36480	107616	4022	532.92
2014-12-30	82006	48640	130646	4372	571.67
2014-12-31	2867	23121	25987	3998	528.86

Table A.2: Daily P&L results for volume order imbalance strategy trading **secondary** futures contract. Pink cells indicate a loss during the session or entire day.

A.3 Final Improved Strategy P&L: Main Contract

Total days with profit: 185
Total days with losses: 58
Average Daily Sharpe Ratio: 0.464
Annualized Sharpe Ratio: 7.243
* all numbers reported in CNY
Strategy Parameters: $k = 5, q = 0.15, w_1 = w_2 = 0.5, L = 5$

Average:	32,701	25,900	58,600	1,798	184.49
Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-01-02	0	0	0	0	0
2014-01-03	23176	5460	28637	1146	98.92
2014-01-06	39037	5168	44205	1446	121.88
2014-01-07	36008	-565	35443	2042	171.38
2014-01-08	34885	11241	46126	1922	162.07
2014-01-09	22887	5567	28454	1638	137.93
2014-01-10	30790	6892	37683	1898	157.79
2014-01-13	16159	7321	23480	1702	140.9
2014-01-14	12188	-4062	8126	1828	151.07
2014-01-15	7723	12808	20530	1104	91.45

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-01-16	9524	6896	16420	810	67.3
2014-01-17	22656	11412	34067	848	69.96
2014-01-20	11749	5363	17113	1148	94.14
2014-01-21	17147	4345	21492	1390	114.84
2014-01-22	28170	11057	39227	1364	114.36
2014-01-23	2867	-4269	-1402	1006	84.71
2014-01-24	13955	8943	22898	846	71.51
2014-01-27	8552	7361	15913	656	54.84
2014-01-28	4983	931	5914	884	73.93
2014-01-29	9571	1003	10574	664	55.73
2014-01-30	5183	10112	15295	552	45.93
2014-02-07	4792	79	4871	498	41.04
2014-02-10	11611	4111	15722	674	56.99
2014-02-11	6352	6898	13250	958	81.65
2014-02-12	16849	-1336	15513	808	69.13
2014-02-13	1233	7266	8499	594	50.81
2014-02-14	4014	1789	5802	436	37.29
2014-02-17	4753	722	5475	256	22.13
2014-02-18	4766	656	5422	380	32.59
2014-02-19	4029	6806	10835	684	58.92
2014-02-20	10437	6566	17003	656	56.58
2014-02-21	10100	5071	15171	680	57.95
2014-02-24	16036	9269	25305	914	75.68
2014-02-25	10684	36302	46986	1182	97.17
2014-02-26	25956	11733	37689	1490	119.96
2014-02-27	11598	21190	32788	1540	124.36
2014-02-28	11801	36872	48673	1654	133.44
2014-03-03	37540	9556	47096	1686	137.12
2014-03-04	24453	15171	39625	1948	157.38
2014-03-05	28527	9417	37944	1800	145.68
2014-03-06	28440	27469	55909	1872	150.65
2014-03-07	29607	9659	39266	1890	153.17
2014-03-10	51722	14819	66541	1980	155.69
2014-03-11	38003	8540	46543	2388	186.09
2014-03-12	65469	21878	87347	2596	202.86
2014-03-13	45000	13235	58236	2372	187.32
2014-03-14	18509	4589	23098	2434	191.41
2014-03-17	55861	-492	55370	2002	159.05
2014-03-18	9043	346	9389	1996	159.16
2014-03-19	22404	8127	30532	1840	144.94
2014-03-20	21833	13697	35530	1468	115.15
2014-03-21	10984	74437	85420	2166	170.2
2014-03-24	32249	20848	53097	2506	201.91
2014-03-25	34040	6108	40148	2796	225.86
2014-03-26	14110	5092	19202	2124	171.59
2014-03-27	14962	41097	56059	1888	152
2014-03-28	44588	20512	65100	1862	150.3
2014-03-31	17891	14329	32219	2318	186.9
2014-04-01	26264	23741	50004	2276	183.78
2014-04-02	16583	1302	17885	2120	172.18
2014-04-03	17150	9272	26421	1604	130.99

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-04-04	13391	6549	19939	1404	114.2
2014-04-08	25564	16327	41892	1582	131.64
2014-04-09	2441	409	2850	1636	137.85
2014-04-10	2906	47903	50809	1828	154.55
2014-04-11	7727	7973	15700	1408	120.1
2014-04-14	7978	3428	11406	1424	121.17
2014-04-15	14539	10835	25374	1034	87.03
2014-04-16	19012	13362	32373	1278	107.23
2014-04-17	-2346	-549	-2895	1140	95.48
2014-04-18	23205	6247	29452	1026	85.24
2014-04-21	32452	10018	42469	1412	117.35
2014-04-22	16036	22105	38142	2322	189.99
2014-04-23	12853	7027	19879	1818	149
2014-04-24	10884	12246	23130	1444	118.35
2014-04-25	28236	16155	44391	1390	113.74
2014-04-28	24376	7336	31712	1460	117.74
2014-04-29	2459	13483	15942	1468	118.29
2014-04-30	1731	-2878	-1147	1028	83.14
2014-05-05	7692	14920	22612	554	44.44
2014-05-06	5424	6591	12015	704	56.93
2014-05-07	6587	7381	13968	1098	88.26
2014-05-08	13201	9315	22516	1054	84.82
2014-05-09	2760	17237	19997	1284	102.81
2014-05-12	31834	11658	43493	1366	111.04
2014-05-13	-260	-590	-850	1260	102.65
2014-05-14	3340	1266	4606	760	61.87
2014-05-15	1303	5551	6855	260	20.93
2014-05-16	957	6808	7765	418	33.48
2014-05-19	11017	1743	12761	606	47.8
2014-05-20	9785	2656	12441	1126	88.99
2014-05-21	19985	6767	26753	1260	99.64
2014-05-22	19849	8478	28326	960	76.77
2014-05-23	3898	15165	19063	1032	82.28
2014-05-26	8152	9774	17927	742	59.77
2014-05-27	1350	5348	6698	676	54.41
2014-05-28	19673	13150	32823	778	62.69
2014-05-29	3531	723	4255	744	60.33
2014-05-30	5761	5097	10858	724	58.51
2014-06-03	4032	1026	5059	404	32.61
2014-06-04	-1621	1097	-524	534	42.52
2014-06-05	1742	7715	9457	478	38.22
2014-06-06	1162	7413	8575	466	37.23
2014-06-09	13968	3572	17540	550	44
2014-06-10	18271	19316	37587	1058	85.06
2014-06-11	4532	-226	4305	1228	99.07
2014-06-12	-1020	-1316	-2337	766	61.78
2014-06-13	7175	7173	14348	472	38.36
2014-06-16	1954	8011	9964	578	47.38
2014-06-17	3868	356	4224	612	49.98
2014-06-18	-3381	3798	417	236	19.22
2014-06-19	3845	6995	10839	330	26.5

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-06-20	5500	6371	11871	338	27.05
2014-06-24	6743	634	7377	402	32.12
2014-06-25	1059	3256	4315	298	23.73
2014-06-26	501	2427	2928	442	35.46
2014-06-27	2601	3195	5796	480	38.52
2014-06-30	12762	1584	14346	456	36.87
2014-07-01	4499	3973	8472	378	30.54
2014-07-02	2938	4303	7241	454	36.69
2014-07-03	5872	7334	13206	376	30.57
2014-07-04	2145	363	2508	274	22.26
2014-07-07	992	1033	2024	230	18.68
2014-07-08	5203	1244	6447	202	16.36
2014-07-09	5158	3277	8435	258	20.83
2014-07-10	-1222	246	-976	192	15.38
2014-07-11	3420	1326	4746	302	24.27
2014-07-14	996	6065	7061	172	13.89
2014-07-15	196	-273	-76	160	12.98
2014-07-16	-4306	208	-4098	134	10.89
2014-07-17	1250	-1033	217	46	3.72
2014-07-18	283	-609	-325	320	25.93
2014-07-21	609	592	1201	252	20.4
2014-07-22	7178	4322	11500	240	19.6
2014-07-23	3668	-1077	2591	324	26.64
2014-07-24	19479	3708	23187	542	45.16
2014-07-25	11141	4652	15793	540	45.54
2014-07-28	71686	18991	90677	1046	91.11
2014-07-29	2498	11720	14218	1562	137.11
2014-07-30	24677	1388	26066	1910	167.27
2014-07-31	13775	1799	15573	1212	106.33
2014-08-01	24675	19439	44114	1180	104.63
2014-08-04	24172	21151	45324	1374	121.98
2014-08-05	21434	14990	36424	1762	156.88
2014-08-06	17277	10603	27879	1890	167.2
2014-08-07	27246	12845	40091	1512	133.44
2014-08-08	7494	7355	14849	1586	139.05
2014-08-11	31675	8156	39831	1500	132.94
2014-08-12	12897	8982	21878	1184	104.51
2014-08-13	10761	11189	21949	1020	90.05
2014-08-14	8152	6199	14350	1100	97.45
2014-08-15	11717	27166	38883	1170	103.79
2014-08-18	25030	9590	34620	1100	98.1
2014-08-19	8592	5768	14360	1192	106.1
2014-08-20	5455	5020	10475	1126	100.02
2014-08-21	5926	4238	10164	812	71.58
2014-08-22	3414	4858	8273	810	71.74
2014-08-25	5675	6419	12094	878	77.53
2014-08-26	14004	6885	20889	870	76.46
2014-08-27	4696	1518	6213	692	60.73
2014-08-28	402	6924	7326	562	49.17
2014-08-29	-3371	2520	-851	418	36.65
2014-09-01	1385	619	2004	366	32.28

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-09-02	1565	16863	18428	456	40.46
2014-09-03	24814	3229	28043	602	54.49
2014-09-04	-1791	9378	7587	804	73.06
2014-09-05	11577	11564	23141	884	81.39
2014-09-09	10665	11664	22329	660	60.86
2014-09-10	9927	-1779	8148	742	67.86
2014-09-11	18732	35283	54015	818	74.92
2014-09-12	11324	4851	16176	1104	100.62
2014-09-15	15726	5323	21050	1132	103.25
2014-09-16	15678	20574	36252	1018	92.64
2014-09-17	7587	4686	12273	918	82.64
2014-09-18	8252	3131	11383	1368	123.68
2014-09-19	8048	17489	25537	1218	110.82
2014-09-22	22193	12319	34512	1186	106.74
2014-09-23	8470	11832	20302	1430	128.89
2014-09-24	22944	30454	53398	1708	156.01
2014-09-25	31533	23611	55144	1694	157.18
2014-09-26	33211	19485	52696	2000	183.82
2014-09-29	8615	4056	12671	1718	158.05
2014-09-30	6769	3388	10157	1350	124.42
2014-10-08	3002	9734	12736	940	86.92
2014-10-09	10202	208	10410	730	67.95
2014-10-10	5117	7900	13017	988	91.52
2014-10-13	19569	-831	18739	962	88.11
2014-10-14	7301	10156	17458	1230	112.81
2014-10-15	16669	12467	29136	1198	109.92
2014-10-16	12353	11051	23404	1056	97.48
2014-10-17	27655	14380	42036	1372	125.82
2014-10-20	17584	-920	16663	1080	99.38
2014-10-21	3112	1969	5081	928	85.3
2014-10-22	5724	4493	10217	622	56.92
2014-10-23	8323	6132	14454	728	66.03
2014-10-24	2407	841	3248	568	51.26
2014-10-27	5530	-1345	4184	596	53.18
2014-10-28	3466	11377	14842	602	54.19
2014-10-29	-557	16476	15919	608	55.71
2014-10-30	10401	9693	20094	514	47.43
2014-10-31	15569	26133	41702	862	80.69
2014-11-03	2632	3544	6176	1006	95.12
2014-11-04	196	552	748	1088	102.46
2014-11-05	3836	-2708	1128	606	57.06
2014-11-06	9834	6285	16120	448	42.1
2014-11-07	6037	23067	29105	736	69.58
2014-11-10	37400	15380	52780	1460	138.7
2014-11-11	37363	45439	82802	2292	220.49
2014-11-12	3453	10249	13702	2140	205.39
2014-11-13	24943	13086	38030	2008	193.85
2014-11-14	12377	7289	19666	1520	146.17
2014-11-17	34068	25959	60027	1610	155.87
2014-11-18	6210	15597	21806	2050	196.07
2014-11-19	-4331	-2486	-6817	1522	145.08

Date	Morning	Afternoon	Total P&L	Trade Volume	Commissions
2014-11-20	12335	5444	17779	852	81.21
2014-11-21	4694	18645	23339	720	69.31
2014-11-24	59283	27491	86774	1402	139.73
2014-11-25	10942	31556	42498	2242	226.11
2014-11-26	58659	27422	86081	2738	280.59
2014-11-27	23011	13739	36750	2668	274.65
2014-11-28	9528	74458	83986	2674	276.97
2014-12-01	50957	32155	83112	2900	306.84
2014-12-02	43445	173840	217285	3818	411.17
2014-12-03	263952	128285	392237	4756	535.72
2014-12-04	211786	191603	403389	6750	780.86
2014-12-05	421575	207395	628969	7720	921.35
2014-12-08	281148	239837	520985	8922	1082.37
2014-12-09	373015	379991	753006	8886	1109.37
2014-12-10	532142	293283	825425	9082	1098.87
2014-12-11	325270	296059	621328	10334	1250.26
2014-12-12	230975	116784	347759	11102	1347
2014-12-15	122493	-7503	114990	10754	1294.65
2014-12-16	235641	98619	334259	9182	1126.8
2014-12-17	56763	16812	73575	8898	1112.02
2014-12-18	201636	246869	448505	6858	897.28
2014-12-19	154247	225034	379280	7048	917
2014-12-22	213463	237297	450760	8812	1149.4
2014-12-23	175206	172424	347631	9186	1175.15
2014-12-24	248019	200450	448470	9590	1176.15
2014-12-25	131819	178632	310451	9612	1188.24
2014-12-26	201691	242689	444379	9064	1162.4
2014-12-29	218471	221582	440053	8856	1174.14
2014-12-30	218631	222219	440850	9534	1246.35
2014-12-31	93627	112697	206323	9718	1284.98

Table A.3: Daily P&L results for final improved strategy with parameters $k = 5, q = 0.15, w_1 = w_2 = 0.5, L = 5$, trading main futures contract. Pink cells indicate a loss during the session or entire day.

Appendix B

Daily P&L Heatmaps for Various Lags

B.1 Lag 2 P&L Heatmap

Trading Threshold (q)

Forecast Window (k)	0.13	0.135	0.14	0.145	0.15	0.155	0.16	0.165	0.17	0.175	0.18	0.185	0.19	0.195	0.2
1	47,074	46,625	45,898	44,863	43,782	42,348	41,720	40,936	40,024	38,845	37,666	37,135	36,323	35,620	34,617
2	54,302	54,693	54,193	54,031	52,868	51,867	50,416	49,104	47,949	46,654	46,082	44,969	44,278	43,022	41,719
3	55,872	56,840	56,950	56,716	56,135	55,137	54,091	52,903	51,828	50,810	49,323	48,204	47,403	46,518	45,880
4	56,285	57,283	57,550	57,635	57,545	56,388	55,513	54,574	53,602	52,566	51,405	49,962	48,738	48,026	47,521
5	55,998	57,170	57,721	57,783	58,035	57,205	55,857	55,052	54,212	53,044	52,219	51,239	50,059	49,033	48,191
6	55,530	56,886	57,706	57,722	58,012	57,420	56,473	55,371	54,495	53,593	52,669	51,548	50,687	49,575	48,822
7	55,542	56,773	57,446	57,924	57,868	57,583	56,461	55,653	54,847	53,719	52,704	52,025	51,014	49,889	49,024
8	55,546	56,772	57,333	57,802	57,936	57,562	56,651	55,759	54,968	54,129	53,028	52,021	51,232	50,076	49,334
9	55,409	56,746	57,436	57,707	57,851	57,554	56,756	55,924	55,059	54,141	53,181	52,173	51,431	50,329	49,513
10	55,095	56,569	57,353	57,795	57,798	57,547	56,953	56,108	55,009	54,270	53,109	52,394	51,405	50,546	49,626
11	54,927	56,434	57,268	57,549	57,854	57,533	57,027	56,265	55,144	54,499	53,296	52,616	51,571	50,697	49,822
12	54,722	56,268	57,005	57,382	57,880	57,745	57,131	56,323	55,389	54,657	53,507	52,640	51,706	50,578	50,084
13	54,590	56,074	57,005	57,357	57,877	57,794	57,115	56,568	55,618	54,658	53,714	52,644	51,698	50,716	50,054
14	54,333	55,935	56,924	57,371	57,816	57,641	57,105	56,618	55,594	54,695	53,847	52,694	51,826	50,978	50,069
15	54,204	55,721	56,827	57,378	57,726	57,551	57,079	56,521	55,669	54,783	53,819	52,820	51,995	51,125	49,919
16	54,136	55,531	56,693	57,291	57,599	57,573	57,045	56,669	55,641	54,897	53,881	52,856	52,015	51,295	50,112
17	54,174	55,524	56,686	57,064	57,486	57,449	57,032	56,783	55,801	54,698	54,005	52,896	52,094	51,351	50,144
18	54,011	55,442	56,474	56,946	57,420	57,326	57,198	56,797	55,883	54,847	54,123	52,967	52,086	51,356	50,161
19	53,885	55,380	56,327	56,927	57,359	57,327	57,082	56,618	55,912	55,000	54,182	53,095	52,038	51,355	50,269
20	53,813	55,203	56,304	56,908	57,287	57,307	57,076	56,491	55,883	55,063	54,324	53,142	52,015	51,330	50,264

Figure B.1: Mean daily profit and loss heatmap for the strategy using the linear model (3.6) with lag $L = 2$, coefficient weights $w_1 = w_2 = 0.5$, over the mesh $q = 0.13, 0.135, ..., 0.2$ and $k = 1, ..., 20$. Darker blue cells denote larger P&L while darker red cells denote lower P&L. The parameters with the largest mean daily P&L is $k = 5$ and $q = 0.15$ indicated by the thick border.

B.2 Lag 3 P&L Heatmap

Trading Threshold (q)

Forecast Window (k)	0.13	0.135	0.14	0.145	0.15	0.155	0.16	0.165	0.17	0.175	0.18	0.185	0.19	0.195	0.2
1	47,471	47,535	46,608	45,471	44,140	43,263	42,450	41,793	40,723	39,781	38,598	37,692	36,972	36,269	35,171
2	54,707	55,214	55,019	54,434	53,772	52,536	51,410	50,159	48,896	47,742	46,972	46,037	45,117	44,014	42,754
3	56,309	57,129	57,684	57,300	56,855	56,273	55,096	54,040	52,943	51,516	50,341	49,358	48,621	47,669	46,845
4	56,129	57,235	57,878	58,113	57,572	56,987	56,344	55,559	54,594	53,338	52,135	51,073	50,108	49,405	48,487
5	55,584	57,103	57,778	57,940	58,174	57,670	56,915	55,979	55,109	53,917	53,082	51,922	51,063	50,195	49,378
6	55,443	56,944	57,612	57,826	58,202	57,744	57,348	56,231	55,336	54,366	53,619	52,489	51,629	50,556	49,833
7	55,379	56,922	57,505	57,983	58,063	57,956	57,476	56,455	55,531	54,492	53,767	52,747	51,775	51,003	50,106
8	55,186	56,800	57,520	57,799	58,169	58,075	57,478	56,630	55,535	54,878	53,736	53,069	52,093	51,170	50,257
9	55,089	56,722	57,555	57,885	58,064	58,085	57,503	56,803	55,694	54,944	53,848	53,003	52,014	51,280	50,616
10	54,962	56,669	57,580	57,711	57,921	58,152	57,694	56,868	55,857	54,912	53,992	52,989	51,998	51,292	50,685
11	54,815	56,561	57,418	57,659	58,072	58,169	57,786	56,835	55,949	54,869	54,196	53,026	52,224	51,462	50,692
12	54,543	56,354	57,318	57,593	57,941	58,192	57,808	57,118	55,971	55,138	54,231	53,240	52,395	51,217	50,825
13	54,371	56,167	57,243	57,483	57,970	58,138	57,804	57,268	56,118	55,161	54,247	53,297	52,537	51,342	51,037
14	54,095	56,024	57,342	57,453	57,923	58,044	57,843	57,295	56,176	55,264	54,231	53,519	52,497	51,495	50,896
15	53,949	55,852	57,192	57,470	57,911	57,987	57,756	57,325	56,330	55,393	54,436	53,565	52,539	51,617	50,821
16	53,693	55,737	57,051	57,544	57,869	57,911	57,851	57,379	56,491	55,356	54,473	53,573	52,520	51,737	50,800
17	53,499	55,552	56,784	57,410	57,819	57,853	57,820	57,413	56,682	55,522	54,509	53,749	52,538	51,788	50,871
18	53,278	55,409	56,729	57,350	57,772	57,789	57,851	57,503	56,674	55,707	54,644	53,657	52,458	51,750	51,022
19	53,092	55,240	56,489	57,297	57,838	57,759	57,651	57,424	56,722	55,709	54,846	53,669	52,562	51,889	50,974
20	52,949	55,205	56,280	57,230	57,666	57,743	57,674	57,399	56,636	55,802	54,890	53,742	52,629	51,901	51,091

Figure B.2: Mean daily profit and loss heatmap for the strategy using the linear model (3.6) with lag $L = 3$, coefficient weights $w_1 = w_2 = 0.5$, over the mesh $q = 0.13, 0.135, ..., 0.2$ and $k = 1, ..., 20$. Darker blue cells denote larger P&L while darker red cells denote lower P&L. The parameters with the largest mean daily P&L is $k = 6$ and $q = 0.15$ indicated by the thick border.

B.3 Lag 4 P&L Heatmap

Trading Threshold (q)

Forecast Window (k)	0.13	0.135	0.14	0.145	0.15	0.155	0.16	0.165	0.17	0.175	0.18	0.185	0.19	0.195	0.2
1	47,696	47,820	46,807	45,747	44,791	43,288	42,695	41,915	40,948	40,079	38,893	38,089	37,242	36,727	35,269
2	54,661	55,527	55,331	54,664	53,859	52,956	51,798	50,356	48,991	48,233	47,461	46,444	45,352	44,045	43,111
3	56,157	57,140	57,725	57,568	57,136	56,392	55,542	54,461	53,018	51,752	50,469	49,777	48,864	48,225	47,198
4	56,083	57,258	57,993	58,206	58,112	57,330	56,438	55,701	54,873	53,615	52,554	51,448	50,643	49,665	49,025
5	55,858	56,958	57,860	58,151	58,291	58,093	57,041	56,129	55,379	54,221	53,421	52,538	51,487	50,574	49,838
6	55,408	57,074	57,792	58,235	58,195	58,167	57,344	56,413	55,503	54,624	53,735	52,932	52,148	51,018	50,179
7	55,338	56,890	57,581	58,240	58,470	58,336	57,586	56,672	55,525	54,682	53,819	53,155	52,439	51,368	50,623
8	55,202	56,990	57,471	58,121	58,474	58,262	57,688	56,769	55,784	54,800	53,952	53,260	52,717	51,861	50,641
9	55,121	56,865	57,504	58,238	58,581	58,330	57,717	56,873	55,952	54,959	54,083	53,424	52,855	52,006	50,798
10	55,110	56,537	57,479	58,141	58,469	58,403	57,624	56,895	56,126	54,976	54,166	53,529	52,912	52,057	51,146
11	54,838	56,353	57,382	58,091	58,335	58,259	57,727	57,016	56,250	55,289	54,278	53,496	52,878	52,176	51,217
12	54,539	56,284	57,300	57,980	58,205	58,265	57,829	57,165	56,456	55,523	54,530	53,631	52,892	52,176	51,331
13	54,217	56,144	56,966	57,891	58,202	58,268	57,946	57,306	56,553	55,762	54,530	53,823	52,946	52,091	51,373
14	53,957	55,962	56,680	57,798	58,134	58,191	57,815	57,391	56,642	55,861	54,731	54,012	52,822	52,143	51,249
15	53,675	55,787	56,677	57,679	58,150	58,251	57,884	57,470	56,668	55,982	54,830	54,040	52,920	52,106	51,273
16	53,307	55,637	56,584	57,575	57,909	58,185	57,840	57,393	56,809	56,008	55,055	54,075	52,924	52,101	51,345
17	53,187	55,436	56,602	57,458	57,740	58,186	57,883	57,323	56,797	56,115	55,207	53,971	52,931	52,227	51,277
18	52,963	55,311	56,480	57,306	57,717	58,064	57,853	57,275	56,878	56,078	55,251	54,064	53,043	52,197	51,340
19	52,791	55,053	56,372	57,144	57,643	58,068	57,803	57,202	56,826	56,172	55,253	54,059	53,043	52,189	51,247
20	52,524	54,910	56,166	57,073	57,505	57,867	57,868	57,318	56,890	56,253	55,254	54,103	53,044	52,353	51,190

Figure B.3: Mean daily profit and loss heatmap for the strategy using the linear model (3.6) with lag $L = 4$, coefficient weights $w_1 = w_2 = 0.5$, over the mesh $q = 0.13, 0.135, ..., 0.2$ and $k = 1, ..., 20$. Darker blue cells denote larger P&L while darker red cells denote lower P&L. The parameters with the largest mean daily P&L is $k = 9$ and $q = 0.15$ indicated by the thick border.

Appendix C

Trading Simulation R Code

```
# function to read in the market data from CSV files
ReadFiles <- function(path, contract) {
  files <- list.files(path = path)
  data <- list()
  for (f in files) {
    key <- substr(f,7,14)
    temp <- read.csv(paste(path,f,sep=''), header=T, stringsAsFactors=F)
    data[[key]] <- temp[which(substr(temp$InstrumentID,1,2)==contract), c(1,3:11,18)]
  }
  save.image(data, file=paste(path,'/DataSet.RData',sep=''))
  return(filepath)
}
```

Listing C.1: **ReadFiles.R**: reads market data CSV files into R dataset

```
GetMainContract <- function(data, open.int = F) {
  if (open.int == T) {
    temp <- data[which(data$SecondOfDay < 33240),]
    temp <- temp[temp$OpenInterest==max(temp$OpenInterest),]
  } else {
    temp <- data[which(data$SecondOfDay == 33300),]
    temp <- temp[temp$Volume==max(temp$Volume),]
  }
  return(head(temp$InstrumentID,1))
}
```

Listing C.2: **GetMainContract.R**: given the market data for a trading day, determines the main contract Instrument ID either based on Open Interest or Volume

```
source(file = 'GetMainContract.R')
BuildLinearData <- function(data, morning = T, open.int = F, delay = 20, lags = 5,
    functions = NULL) {
  library(zoo)
  library(TTR)
  # declare constants
  day.start <- 33300 # 9:15
  AM.start <- 33300 # 9:15
  AM.open <- 33360 # 9:16 - trade open
  AM.close <- 40800 # 11:20 - trade close
  AM.end <- 41280 # 11:28

  PM.start <- 46800 # 13:00
  PM.open <- 46860 # 13:01 - trade open
  PM.close <- 54000 # 15:00 - trade close
  PM.end <- 54780 # 15:13
  start.time <- ifelse(morning, AM.start, PM.start) # - data start
```

```r
open.time <- ifelse(morning, AM.open, PM.open)     # - trade open
close.time <- ifelse(morning, AM.close, PM.close) # - trade close
end.time <- ifelse(morning, AM.end, PM.end)        # - data end

# get main contract
instrument <- GetMainContract(data, open.int)

ind <- which(data$InstrumentID==instrument & data$SecondOfDay >= start.time & data$
    SecondOfDay < end.time)
main.data <- data[ind,]
n <- nrow(main.data)

time.secs <- main.data$SecondOfDay + main.data$UpdateMillisec/1000
ind.open <- head(which(time.secs>=open.time),1)
ind.close <- head(which(time.secs>=close.time),1)

# calculate variables
mid.price <- (main.data$BidPrice1 + main.data$AskPrice1)/2
spread <- main.data$AskPrice1-main.data$BidPrice1

OIR.array <- (main.data$BidVolume1 - main.data$AskVolume1) / (main.data$BidVolume1
    + main.data$AskVolume1)
dBid.price <- c(0,diff(main.data$BidPrice1))
dAsk.price <- c(0,diff(main.data$AskPrice1))

## build order imbalance signal according to Spec
bid.CV <- (main.data$BidVolume1 - ifelse(dBid.price==0,c(0,main.data$BidVolume1[-n
    ]),rep(0,n)))*as.integer(dBid.price>=0)
ask.CV <- (main.data$AskVolume1 - ifelse(dAsk.price==0,c(0,main.data$AskVolume1[-n
    ]),rep(0,n)))*as.integer(dAsk.price<=0)
VOI.array <- bid.CV - ask.CV

dVol <- c(NA,diff(main.data$Volume))
dTO <- c(NA,diff(main.data$Turnover))
AvgTrade.price <- dTO / dVol / 300
AvgTrade.price[which(is.nan(AvgTrade.price))] <- NA
AvgTrade.price <- na.locf(na.locf(AvgTrade.price, na.rm=F), fromLast=T)
MPB.array <- (AvgTrade.price - c(mid.price[1], rollmean(mid.price, k=2)))


k <- delay
p <- lags
new.ind <- (p+1):(n-k)

## arithmetic average of future k midprices minus current midprice
if (k > 0) {
  library(zoo)
  fpc <- rollmean(mid.price, k=k)[-1] - mid.price[1:(n-k)]
  dMid.Response <- c(fpc, rep(NA,k))
} else {
  dMid.Response <- rep(0,n)
}


# build VOI, dMid, OIR - first p entries are useless
VOI <- cbind(VOI.array)
OIR <- cbind(OIR.array)
MPB <- cbind(MPB.array)
if (p > 0) {
  for (j in 1:p) {
    VOI <- cbind(VOI, c(rep(NA,j), VOI.array[1:(n-j)]))
    OIR <- cbind(OIR, c(rep(NA,j), OIR.array[1:(n-j)]))
    MPB <- cbind(MPB, c(rep(NA,j), MPB.array[1:(n-j)]))
  }
}


# trim the variables
```

```r
  dMid.Response <- dMid.Response[new.ind]
  VOI <- VOI[new.ind,,drop=FALSE]
  OIR <- OIR[new.ind,,drop=FALSE]
  MPB <- MPB[new.ind,,drop=FALSE]

  colnames(VOI) <- paste('VOI.t',seq(0,p),sep='')
  colnames(OIR) <- paste('OIR.t',seq(0,p),sep='')
  colnames(MPB) <- paste('MPB.t',seq(0,p),sep='')

  # trim the other supporting data
  mid.price <- mid.price[new.ind]
  spread <- spread[new.ind]
  AvgTrade.price <- AvgTrade.price[new.ind]
  main.data <- main.data[new.ind,]
  time.secs <- time.secs[new.ind]

  ind.open <- ind.open - p
  ind.close <- ind.close - p

  # return an R object
  value <- {}
  value$data <- main.data
  value$dMid.Response <- dMid.Response
  value$VOI <- VOI
  value$OIR <- OIR
  value$MPB <- MPB

  value$time.secs <- time.secs
  value$ind.open <- ind.open
  value$ind.close <- ind.close

  value$mid.price <- mid.price
  value$spread <- spread
  value$AvgTrade.price <- AvgTrade.price

  return(value)
}
```

Listing C.3: **BuildLinearData.R**: given the market data for a trading session, builds the variables to be used in a linear model for the trading strategy

```r
source(file = 'BuildLinearData.R')
BuildLinearModel <- function(key, data, full.day = T, morning = T, open.int = F,
    delay = 20, lags = 5, strategy = '', functions = NULL) {
  # check if we need a full-day linear model or for a single trading session
  if (full.day == T) {
    morning.data <- BuildLinearData(data, morning=T, open.int=open.int, delay=delay,
        lags=lags, functions=functions)
    evening.data <- BuildLinearData(data, morning=F, open.int=open.int, delay=delay,
        lags=lags, functions=functions)
    dMid.Response <- c(morning.data$dMid.Response, evening.data$dMid.Response)
    VOI <- rbind(morning.data$VOI, evening.data$VOI)
    OIR <- rbind(morning.data$OIR, evening.data$OIR)
    time.secs <- c(morning.data$time.secs, evening.data$time.secs)
    mid.price <- c(morning.data$mid.price, evening.data$mid.price)
    spread <- c(morning.data$spread,evening.data$spread)
    AvgTrade.price <- c(morning.data$AvgTrade.price, evening.data$AvgTrade.price)
    MPB <- rbind(morning.data$MPB, evening.data$MPB)
    trading.data <- rbind(morning.data$data, evening.data$data)
  } else {
    trading.data <- BuildLinearData(data, morning = morning, open.int = open.int,
        delay = delay, lags = lags, functions=functions)
    dMid.Response <- trading.data$dMid.Response
    VOI <- trading.data$VOI
    OIR <- trading.data$OIR
    time.secs <- trading.data$time.secs
    mid.price <- trading.data$mid.price
```

```
    spread <- trading.data$spread
    AvgTrade.price <- trading.data$AvgTrade.price
    MPB <- trading.data$MPB
    trading.data <- trading.data$data
  }


  ## build the features matrix (x-variable) based on strategy
  ## transform the variables if necessary
  identity <- function(x) x
  inverse <- function(x) 1/x
  f.VOI <- if(is.null(functions[['VOI']])) identity else functions[['VOI']]
  f.OIR <- if(is.null(functions[['OIR']])) identity else functions[['OIR']]

  ## build the explanatory variables
  x <- list()
  x[['A']] <- data.frame(y=dMid.Response, VOI=f.VOI(VOI))
  x[['B']] <- data.frame(y=dMid.Response, VOI=f.VOI(VOI)/spread, OIR=f.OIR(OIR)/
     spread, MPB=MPB[,1]/spread)

  value <- {}
  # build the linear model using OLS
  if (strategy != '') {
    s <- strategy
    value$model <- lm(y ~ ., data=x[[s]])
  }

  ## return values
  value$dMid.Response <- dMid.Response ## y-value
  value$VOI <- VOI
  value$OIR <- OIR
  value$spread <- spread
  value$y <- dMid.Response
  value$x <- x
  value$data <- trading.data
  value$AvgTrade.price <- AvgTrade.price
  value$mid.price <- mid.price
  value$MPB <- MPB
  value$time.secs <- time.secs


  return(value)
}
```

Listing C.4: **BuildLinearModel.R**: given the market data for a trading session, builds the linear model to be used by the strategy

```
#####################################################
##  LINEAR MODEL STRATEGY:
##  BUY SIGNAL (at t)
##    * E[FPC(t)] >= 0.2
##
##  SELL SIGNAL (at t)
##    * E[FPC(t)] <= -0.2
##
##  if signal hits, buy or sell maximum position
#####################################################
source(file = 'GetMainContract.R')
source(file = 'BuildLinearData.R')
LinearStrategy <- function(data, coefs, lags, strategy = 'A', threshold = 0.2,
    morning = T, open.int = F, trade.at.mid = F, functions = NULL) {

  ## get all the market data (this would be a data-stream in a real-time system)
  TR.COST <- 2.5*1e-5
  value <- BuildLinearData(data, morning = morning, open.int = open.int, delay = 0,
     lags = lags)
  main.data <- value$data
```

```
  n <- nrow(main.data)

  mid.price <- value$mid.price
  spread <- value$spread
  time.secs <- value$time.secs
  ind.open <- value$ind.open
  ind.close <- value$ind.close

  own <- F
  pos <- 0
  strat <- rep(0,n)
  realized.pnl <- rep(NA,n)
  total.trade.pnl <- c()
  returns <- c()
  pnl <- 0
  trade.costs <- 0
  buy.price <- 0
  sell.price <- 0
  entry <- 0
  trade.volume <- 0
  sharpes <- c()

  # get the vector of bid/ask prices (this will be scalar in data stream)
  ask <- if(trade.at.mid) mid.price else main.data$AskPrice1
  bid <- if(trade.at.mid) mid.price else main.data$BidPrice1

  # Set the x-values to be used in prediction depending on strategy
  # these would be scalar in a data stream
  VOI <- value$VOI
  OIR <- value$OIR
  MPB <- value$MPB
  identity <- function(x) x
  f.VOI <- if(is.null(functions[['VOI']])) identity else functions[['VOI']]
  f.OIR <- if(is.null(functions[['OIR']])) identity else functions[['OIR']]

  x <- cbind(rep(1,n))
  if (strategy == 'A') {
    x <- cbind(x, f.VOI(VOI))
  } else if (strategy == 'B') {
    x <- cbind(x, f.VOI(VOI) / spread, f.OIR(OIR) / spread, MPB[,1] / spread)
  } else {
    stop(paste('Missing Linear Strategy:', strategy))
  }


  # this is where we assume we get a data stream instead of looping through the dataset
  # multiply the coefficients with the factors and check if it's above/below threshold
  # and trade if the signal is good

  # in an actual trading system, the decision would be calculated by a strategy engine
  # having the real-time data fed into the engine via a data stream
  # but in this simulation, we just assume we have the full dataset and the
  # strategy engine is the coefficient multiplication on the next line
  efpc.vec <- rowSums(x * matrix(rep(coefs,n),byrow=T, nrow=n))
  # each k = 500ms;
  for (k in trade.ind) {
    efpc <- efpc.vec[k]

    ## check if we are within trading hours
    if(k >= ind.open & k < ind.close & own == F & efpc >= threshold) {
      ## BUY to OPEN
      strat[k] <- 1
      own = T
      pos <- 1
      buy.price <- ask[k]
```

```
      entry <- k
      tc <- buy.price * TR.COST
      trade.costs <- trade.costs + tc
      trade.volume <- trade.volume + 1
    } else if (k >= ind.open & k < ind.close & own == F & efpc <= -threshold) {
      ## SELL to OPEN
      strat[k] <- -1
      own = T
      pos <- -1
      sell.price <- bid[k]
      entry <- k
      tc <- sell.price * TR.COST
      trade.costs <- trade.costs + tc
      trade.volume <- trade.volume + 1
    } else if (own == T & pos == 1 & efpc <= -threshold) {
      ## SELL to CLOSE
      strat[k] <- -1
      own <- F
      pos <- 0
      sell.price <- bid[k]
      tc <- tc + sell.price * TR.COST
      trade.costs <- trade.costs + tc
      trade.pnl <- sell.price-buy.price - tc
      pnl <- pnl + trade.pnl
      trade.volume <- trade.volume + 1
      total.trade.pnl <- c(total.trade.pnl, trade.pnl)

      if (k >= ind.open & k < ind.close) {
        ## SELL to OPEN
        strat[k] <- -2
        own <- T
        pos <- -1
        sell.price <- bid[k]
        entry <- k
        tc <- sell.price * TR.COST
        trade.costs <- trade.costs + tc
        trade.volume <- trade.volume + 1
      }
    } else if (own == T & pos == -1 & efpc >= threshold) {
      ## BUY to CLOSE
      strat[k] <- 1
      own = F
      pos <- 0
      buy.price <- ask[k]
      tc <- tc + buy.price * TR.COST
      trade.costs <- trade.costs + tc
      trade.pnl <- sell.price-buy.price - tc
      pnl <- pnl + trade.pnl
      trade.volume <- trade.volume + 1
      total.trade.pnl <- c(total.trade.pnl, trade.pnl)

      if (k >= ind.open & k < ind.close) {
        ## BUY to OPEN
        strat[k] <- 2
        own <- T
        pos <- 1
        buy.price <- ask[k]
        entry <- k
        tc <- buy.price * TR.COST
        trade.costs <- trade.costs + tc
        trade.volume <- trade.volume + 1
      }
    }
    realized.pnl[k] <- pnl
  }

  # check if we have a left-over position at end-of-day and close it
  if (sum(strat) == 1) {
```

```
    if (strat[n] == 1) {
      strat[n] <- 0
      trade.volume <- trade.volume - 1
    } else {
      strat[n] <- -1
      sell.price <- bid[n]
      tc <- tc + sell.price * TR.COST
      trade.costs <- trade.costs + tc
      trade.pnl <- sell.price-buy.price - tc
      pnl <- pnl + trade.pnl
      realized.pnl[n] <- pnl
      total.trade.pnl <- c(total.trade.pnl, trade.pnl)
      trade.volume <- trade.volume + 1
    }
  } else if (sum(strat)==-1) {
    if (strat[n] == -1) {
      strat[n] <- 0
      trade.volume <- trade.volume - 1
    } else {
      strat[n] <- 1
      buy.price <- ask[n]
      tc <- tc + buy.price * TR.COST
      trade.costs <- trade.costs + tc
      trade.pnl <- (sell.price-buy.price) - tc
      pnl <- pnl + trade.pnl
      realized.pnl[n] <- pnl
      total.trade.pnl <- c(total.trade.pnl, trade.pnl)
      trade.volume <- trade.volume + 1
    }
  }

  # return stats
  realized.pnl <- na.locf(c(0,realized.pnl))[-1]

  value <- {}
  value$time <- time.secs
  value$pnl <- realized.pnl
  value$strategy <- strat
  value$trade.volume <- trade.volume
  value$trade.pnl <- total.trade.pnl
  value$trade.costs <- trade.costs
  return(value)
}
```

Listing C.5: **LinearStrategy.R**: the trading strategy given a linear model to forecast price changes

```
load(file = 'DataSet.RData')

source('BuildLinearModel.R')
source('LinearStrategy.R')

#### AVERAGED LAG LINEAR STRATEGY ####

## set trading and model parameters
threshold <- 0.2
period <- 20
lags <- 5
strategy <- 'A'
coefs <- c()

## build the linear models and store their coefficients
for (i in 1:length(data)) {
  key <- names(data)[i]
  # full-day coefficients
  value <- BuildLinearModel(key, data[[i]], full.day = T, delay = period, lags = lags
      , strategy = strategy)
```

```r
  model <- value$model
  coefs <- rbind(coefs, model$coefficients)
  print(names(data)[i])
}

## set the lagged coefficient weights
coef.weights <- c(1)
trade.volume <- c()
trade.costs <- c()

pnl.name <- paste('pnl-',threshold,'-',strategy,period,'F-lag',lags,sep='')
assign(pnl.name, matrix(nrow=length(data), ncol=2))
pnl.matrix <- get(pnl.name)
pnl.matrix[1,] <- 0
trade.pnl <- c()

## apply the trading strategy to each trading day using historical linear model coefficients
for (i in 1:length(data)) {
  key <- names(data)[i]

  if (i > 1) {
    coef <- 0
    w <- coef.weights[1:min(length(coef.weights),i-1)]
    w <- w/sum(w)
    for (j in 1:length(w)) {
      coef <- coef + coefs[i-j,] * w[j]
    }

    # morning trading using the weighted coefficients from T-1, T-2,...
    strat <- LinearStrategy(data[[i]], coef, lags=lags, strategy=strategy, morning = T, threshold=threshold)
    pnl.matrix[i,1] <- tail(strat$pnl,1)
    trade.pnl <- c(trade.pnl, strat$trade.pnl)
    tv <- strat$trade.volume
    tc <- strat$trade.costs

    # afternoon trading using the weighted coefficients from T-1, T-2,...
    strat <- LinearStrategy(data[[i]], coef, lags=lags, strategy=strategy, morning = F, threshold=threshold)
    pnl.matrix[i,2] <- tail(strat$pnl,1)
    trade.pnl <- c(trade.pnl, strat$trade.pnl)

    tv <- tv + strat$trade.volume
    trade.volume <- c(trade.volume, tv)
    tc <- tc + strat$trade.costs
    trade.costs <- c(trade.costs, tc)
  }
  print(paste(key,strategy,period,threshold,'P&L =',pnl.matrix[i,1],pnl.matrix[i,2],'Total =',sum(pnl.matrix[1:i,])))
}
assign(pnl.name, pnl.matrix)
sharpe.ratio <- mean(rowSums(pnl.matrix)) * sqrt(nrow(pnl.matrix)) / sd(rowSums(pnl.matrix))
write.table(pnl.matrix, paste(pnl.name,'.txt',sep=''), sep='\t', row.names=F, col.names=F)
```

Listing C.6: **RunStrategy.R**: script to run the linear trading strategy