CAPTURING THE ORDER IMBALANCE WITH HIDDEN MARKOV MODEL: A CASE OF SET50 AND KOSPI50

BY

MR. PO-LIN WU

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE PROGRAM IN FINANCE (INTERNATIONAL PROGRAM)

FACULTY OF COMMERCE AND ACCOUNTANCY
THAMMASAT UNIVERSITY

ACADEMIC YEAR 2016

CAPTURING THE ORDER IMBALANCE WITH HIDDEN MARKOV MODEL: A CASE OF SET50 AND KOSPI50

BY

MR. PO-LIN WU

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE PROGRAM IN FINANCE (INTERNATIONAL PROGRAM) FACULTY OF COMMERCE AND ACCOUNTANCY THAMMASAT UNIVERSITY ACADEMIC YEAR 2016

THAMMASAT UNIVERSITY
FACULTY OF COMMERCE AND ACCOUNTANCY

INDEPENDENT STUDY

BY

MR. PO-LIN WU

ENTITLED

CAPTURING THE ORDER IMBALANCE WITH HIDDEN MARKOV MODEL:
A CASE OF SET50 AND KOSPI50

was approved as partial fulfillment of the requirements for
the degree of Master of Science (Finance)

on 01 MAY 2017

Chairman A. Charoenrook

(Assistant Professor Anchada Charoenrook, Ph.D.)

Member and Advisor Wasin S.

(Wasin Siwasarit, Ph.D.)

Dean P. Udorn

(Associate Professor Pipop Udorn, Ph.D.)

Ref. code: 25595802042019LHA

Independent Study Title	CAPTURING THE ORDER IMBALANCE WITH HIDDEN MARKOV MODEL: A CASE OF SET50 AND KOSPI50
Author	Mr. Po-Lin Wu
Degree	Master of Science (Finance)
Major Field/Faculty/University	Master of Science Program in Finance (International Program) Faculty of Commerce and Accountancy Thammasat University
Independent Study Advisor	Wasin Siwarsarit, Ph.D.
Academic Year	2016

ABSTRACT

Based on the empirical evidence of the recent strand of the literature, Market Efficiency creation process is not instantaneous, but rather attains over short-horizon of time. With the low liquidity market, the price movement of financial assets can be predicted by order imbalance indicators. In contrast, in a more liquidity market, the predictability of return is significantly decreased. In this study, we implement one of the well-known machine learning models for pattern recognition known as the Hidden Markov Model (HMM) with order imbalance to forecast the price movement of selected stocks in markets with different levels of liquidity which are the Stock Exchange of Thailand (SET) and Korea Exchange (KRX). As the consequence, we can create an algorithmic trading strategy based on the states of risky assets captured by the models. The result is consistent with the previous literature that both the predictability of the models and the profitability of the strategy diminish as the frequency decreases and market liquidity increases. Remarkably, our model in the market with lower liquidity is able to generate signal that achieves average hit ratio of 83.38% in predicting the risky assets' positive price movement at frequency of 5 minutes.

Keywords: Algorithmic trading, HMM, market efficiency, liquidity, order imbalance

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my Advisor, Wasin Siwasarit, Ph.D. for his patience and continuous guidance throughout this research. Without his support, this study would not have been successful and accomplished. I would also like to thank the committee member Assistant Professor Anchada Charoenrook, Ph.D. for her precious inputs and advises.

Finally, I would like to thank my family for their encouragement and support in the last 2 years.

Mr. Po-Lin Wu

TABLE OF CONTENTS

	Page
ABSTRACT	(1)
ACKNOWLEDGEMENTS	(2)
LIST OF TABLES	(6)
LIST OF FIGURES	(10)
LIST OF ABBREVIATIONS	(11)
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 Literature Review	3
2.1 Application of the Hidden Markov Based model in Finance	3
2.2 Hidden Markov Model in Financial Time Series Forecasting	4
2.3 Other machine Learning Model applied in Financial Time Series Forecasting	5
2.4 Order Imbalances and Stock Return	5
CHAPTER 3 Theoretical Framework	7
3.1 Efficient Market Hypothesis	7
3.2 Market Efficiency, Order Imbalance and Market Liqudity	7

CHAPTER 4 Methodology Framework 9

4.1 Microstructure of SET and KRX 9
4.1.1 Daily Price Limitation 9
4.1.2 Tick Size 9
4.2 Order Imbalance 10
4.3 Data Quantization 11
4.4 Hidden Markov Model 12
4.4.1 Three Fundamental Problems of Hidden Markov Model 13
4.4.2 Number of States in the Hidden Markov Model 18
4.4.3 Generate Trading Signals 18
4.4.3.1 Discrete Case 18
4.4.3.2 Continuous Case 19
4.4.3.2.1 Gaussian Mixture Model 20
4.4.3.2.2 HMM with the Gaussian Mixtures 21
4.4.3.2.3 Generating Trading Signals 22
4.5 Trading Strategy 23
4.6 Performance Measurement 24
4.6.1 Benchmark 24
4.6.2 Hit Ratio 24
4.6.3 Sharpe Ratio 25
4.6.4 Jenson's Alpha 25
4.7 Data 26

CHAPTER 5 Result and Discussion 29

5.1 Predictability of the Discrete Model 29
5.1.1 Predictability of the Discrete HMM for Selected Stocks in SET50 29
5.1.2 Predictability of the Discrete HMM for Selected Stocks in KOSPI50 30

5.2 Predictability of the Continuous Model 34

5.3 Performance of Trading Strategy 37

5.3.1 Profitability of the Discrete Model in both Thai and Korean Stock Market 37

5.3.2 Profitability of the Continuous Model in both Thai and Korean Stock Market 38

CHAPTER 6 Conclusion and Recommendation 47

6.1 Performance of the Discrete and Continuous Models 47

6.2 Implication on the Performance of the Models in Different Frequencies 48

6.3 Implication on the Performance of the Models in Different Markets 49

6.4 Recommendation for Further Study 49

REFERENCES 51

APPENDICES

APPENDIX A 54

APPENDIX B 60

APPENDIX C 63

APPENDIX D 69

APPENDIX E 73

APPENDIX F 80

BIOGRAPHY 85

LIST OF TABLES

Tables		Page
4.1	Tick Size of SET Listing	9
4.2	Tick Size of KRX Listing	10
4.3	Data Quantization for the discrete Hidden Markov Model	11
4.4	25% and 75% Percentile of Order Imbalance Ratio	12
4.5	The listed stocks selected from SET50 for study	27
4.6	The listed stocks selected from KOSPI50 for study	27
5.1	Hit Ratio of discrete HMM for selected stocks in SET50	32
5.2	Hit Ratio of discrete HMM for selected stocks in KOSPI50	33
5.3	Hit Ratio of continuous HMM for selected stocks in SET50	35
5.4	Hit Ratio of continuous HMM for selected stocks In KOSPI50	36
5.5	Sharpe Ratio of Discrete HMM in SET50	39
5.6	Sharpe Ratio of Discrete HMM in KOSPI50	40
5.7	Sharpe Ratio of Continuous HMM in SET50	41
5.8	Sharpe Ratio of Continuous HMM in KOSPI50	42
5.9	Jenson’s Alpha (Discrete HMM, SET50, 0.05% transaction cost)	43
5.10	Jenson’s Alpha (Discrete HMM, KOSPI50, 0.05% transaction cost)	44
5.11	Jenson’s Alpha (Continuous HMM, SET50, 0.05% transaction cost)	45
5.12	Jenson’s Alpha (Continuous HMM, KOSPI50, 0.05% transaction cost)	46
A.1	Summary statistics of SET50: return of 5 Minute data	54
A.2	Summary statistics of SET50: Order Imbalance ratio of 5 Minute data	54
A.3	Summary statistics of SET50: Correlation between return and OIR of 5 minute data	54
A.4	Summary statistics of SET50: return of 10 Minute data	55
A.5	Summary statistics of SET50: Order Imbalance ratio of 10 Minute data	55
A.6	Summary statistics of SET50: Correlation between return and OIR of 10 minute data	55
A.7	Summary statistics of SET50: return of 30 Minute data	56
A.8	Summary statistics of SET50: Order Imbalance ratio of 30 Minute data	56

A.9 Summary statistics of SET50: Correlation between return and OIR of 30 minute data 56
A.10 Summary statistics of KOSPI50: return of 5 Minute data 57
A.11 Summary statistics of KOSPI50: Order Imbalance ratio of 5 Minute data 57
A.12 Summary statistics of KOSPI50: Correlation between return and OIR of 5 minute data 57
A.13 Summary statistics of KOSPI50: return of 10 Minute data 58
A.14 Summary statistics of KOSPI50: Order Imbalance ratio of 10 Minute data 58
A.15 Summary statistics of KOSPI50: Correlation between return and OIR of 10 minute data 58
A.16 Summary statistics of KOSPI50: return of 30 Minute data 59
A.17 Summary statistics of KOSPI50: Order Imbalance ratio of 30 Minute data 59
A.18 Summary statistics of KOSPI50: Correlation between return and OIR of 10 minute data 59
B.1 Normality Test on SET50 5 Minute Data 60
B.2 Normality Test on SET50 10 Minute Data 60
B.3 Normality Test on SET50 30 Minute Data 61
B.4 Normality Test on KOSPI50 5 Minute Data 61
B.5 Normality Test on KOSPI50 10 Minute Data 62
B.6 Normality Test on KOSPI50 30 Minute Data 62
C.1 GMM Fitting: SET50 5 minute data 63
C.2 GMM Fitting: SET50 10 minute data 64
C.3 GMM Fitting: SET50 30 minute data 65
C.4 GMM Fitting: KOSPI50 5 minute data 66
C.5 GMM Fitting: KOSPI50 10 minute data 67
C.6 GMM Fitting: KOSPI50 30 minute data 68
D.1 Sharpe ratio of trading of the Discrete Hidden Markov Model in SET50 (assume no transaction cost) 69
D.2 Sharpe ratio of trading of the Discrete Hidden Markov Model in SET50 (assume 0.05% transaction cost) 69
D.3 Sharpe ratio of trading of the Discrete Hidden Markov Model in SET50 (assume 0.1% transaction cost) 69

D.4 Sharpe ratio of trading of the Continuous Hidden Markov Model in SET50 (assume no transaction cost) 70
D.5 Sharpe ratio of trading of the Continuous Hidden Markov Model in SET50 (assume 0.05% transaction cost) 70
D.6 Sharpe ratio of trading of the Continuous Hidden Markov Model in SET50 (assume 0.1% transaction cost) 70
D.7 Sharpe ratio of trading of the Discrete Hidden Markov Model in KOSPI50 (assume no transaction cost) 71
D.8 Sharpe ratio of trading of the Discrete Hidden Markov Model in KOSPI50 (assume 0.05% transaction cost) 71
D.9 Sharpe ratio of trading of the Discrete Hidden Markov Model in KOSPI50 (assume 0.1% transaction cost) 71
D.10 Sharpe ratio of trading of the Continuous Hidden Markov Model in KOSPI50 (assume no transaction cost) 72
D.11 Sharpe ratio of trading of the Continuous Hidden Markov Model in KOSPI50 (assume 0.05% transaction cost) 72
D.12 Sharpe ratio of trading of the Continuous Hidden Markov Model in KOSPI50 (assume 0.1% transaction cost) 72
E.1 Jenson’s Alpha (Discrete case, SET50, no transaction cost) 73
E.2 Jenson’s Alpha (Discrete case, SET50, 0.05% transaction cost) 74
E.3 Jenson’s Alpha (Discrete case, SET50, 0.01% transaction cost) 74
E.4 Jenson’s Alpha (Continuous case, SET50, no transaction cost) 75
E.5 Jenson’s Alpha (Continuous case, SET50, 0.05% transaction cost) 75
E.6 Jenson’s Alpha (Continuous case, SET50, 0.1% transaction cost) 76
E.7 Jenson’s Alpha (Discrete case, KOSPI50, no transaction cost) 76
E.8 Jenson’s Alpha (Discrete case, KOSPI50, 0.05% transaction cost) 77
E.9 Jenson’s Alpha (Discrete case, KOSPI50, 0.1% transaction cost) 77
E.10 Jenson’s Alpha (Continuous case, KOSPI50, no transaction cost) 78
E.11 Jenson’s Alpha (Continuous case, KOSPI50, 0.05% transaction cost) 78
E.12 Jenson’s Alpha (Continuous case, KOSPI50, 0.1% transaction cost) 79
F.1 Total number of signals that resulted in positive or negative price movement (Discrete case, SET50) 80

F.2 Total number of signals that resulted in positive price movement (Discrete case, SET50) 81
F.3 Total number of signals that resulted in positive or negative price movement (Continuous case, SET50) 81
F.4 Total number of signals that resulted in positive price movement (Continuous, SET50) 82
F.5 Total number of signals that resulted in positive or negative price movement (Discrete case, KOSPI50) 82
F.6 Total number of signals that resulted in positive price movement (Discrete case, KOSPI50) 83
F.7 Total number of signals that resulted in positive or negative price movement (Continuous case, KOSPI50) 83
F.8 Total number of signals that resulted in positive price movement (Continuous case, KOSPI50) 84

LIST OF FIGURES

Figures		Page
4.1	A graph of a simple Hidden Markov Model	12
4.2	Example of a 3 Components Gaussian Mixture Model	21
4.3	2 states HMM with Gaussian Mixture Model	22

LIST OF ABBREVIATIONS

Symbols/Abbreviations	Terms
HMM	Hidden Markov Model
SET	Stock Exchange of Thailand
KOSPI	Korean Composite Stock Price Index
KRX	Korea Exchange
NYSE	New York Stock Exchange
OIR	Order Imbalance Ratio
S&P	Standard and Poor’s

CHAPTER 1
INTRODUCTION

The Efficient Market Hypothesis defined by Fama (1970) states that the asset price should fully reflect all available information; thus, the asset return is not predictable and passive trading is always the optimal trading strategy. However, the empirical evidence from the study by Chordia T. and Roll R. (2005) based on 150 stocks listed on NYSE during year 1996, 1999 and 2002 shows that the market is not strong-form efficient; the future return of selected assets was predictable over the interval of 5 to 30 minutes by using the order imbalance.

The efficiency creating process is also affect by the market liquidity. The previous literature also shows that the predictability of order imbalance is linked to the liquidity of market. The research by Chordia T. et al (2008) provided strong evidence that at a more liquid regime, the predictability of asset tends to disappear due to investors taking advantage of low bid-ask spread.

Based on previous literatures, to beat the market, we should focus on the intra-day frequency, in which the market efficiency is possibly not attained. In addition to the literature of market efficiency, the quantitative hedge fund firms, such as Renaissance Technologies and Two Sigma in the US, were able to outperform the market by utilizing systematic and algorithmic trading and have been actively hiring professionals from field of information theory, which is a field that specializes in symbol and pattern recognition. Their success shows that, even in a market that is highly liquid, the market is still predictable at very high frequency. Therefore, the technology or the models that are utilized in the quantitative trading should be further studied.

However, the algorithms utilized by the top quantitative firms are mostly kept as firm trading secrets and are not likely to be revealed to the public. In the past decade, machine learning models such as, Artificial Neural Network (ANN), Hidden Markov Model (HMM), Fuzzy Logic and Support Vector Machine (SVM), have been proposed in the literature as a way to obtain more accurate forecast.

Hidden Markov Model is a stochastic model and is considered a Bayesian Interference Network. It is often used in analyzing and predicting time series data of different fields such as information theory, weather prediction, and Bioinformatics. In the quantitative finance field, the model has been proposed in the literature to forecast various financial time series. For example, Patrik, I. & Conny, J. (2008) proposed an algorithm to automatic foreign exchange trading base on prediction from Hidden Markov Model, Tenyakov, A. (2014) proposed a Hidden Markov Model based model for filtering and forecasting commodity future prices. The model is versatile in the sense that it is able to take multiple factors into account, such as news, investor's behaviors, and other macroeconomic factors.

This study aims to introduce the Hidden Markov Model and test its prediction ability in forecasting intra-days price movement of selected stocks in the SET50 index and the KOSPI 50 Index. We address the empirical evidence of return predictability by building a trading strategy and back-tested with the inclusion of transaction cost based on the patterns we discover with the proposed model. We also compare the performance of our model with the conventional buy and hold strategy in two different markets.

This paper contribute to 1) the advancement of algorithmic trading in Thailand 2) formulation of trading strategies for institutional or individual traders 3) study of the applicability of machine learning model in the Thai and Korean capital market 4) study of market efficiency in countries with different stock market liquidity at intraday frequency.

The paper is organized as follows. The next chapter documents review on literature. Chapter 3 describes the related theoretical framework of this study. Chapter 4 presents the methodologies of this study. Chapter 5 reports the result on both predictability and profitability of the HMM model. Last chapter contains the discussion and further recommendation of this study.

CHAPTER 2
REVIEW OF LITERATURE

2.1 Application of Hidden Markov based Model in Finance

The Hidden Markov Model is a popular model in engineering fields such as, speech recognition, handwriting and bioinformatics. In field of quantitative finance, factors such as, price and economic indicators are often taken as inputs in the Hidden Markov based model to evaluate the dynamics or hidden variables of observed information. The accurate estimation of these variables would have fair impact on several topics in the finance field. For example, the Hidden Markov model is applied in the studies of credit quality; the study by Korolkiewicz & Elliott (2008) applied the Hidden Markov Model with input of Standard & Poor's credit rating data to estimate the dynamic of credit rating. The result is a transition matrix that represents the probability of change of rating from initial to another. Another earlier research by Giacomo, Mark, and Crowder (2005) applied the Hidden Markov Model with the credit rating data of US bond issues in consumer, energy, media and transport sectors to estimate the sequence of hidden risk states. In addition, Haipeng, Ning, & Ying (2012) extends the study by applying the Hidden Markov Model to estimate the time-varying rating transition and capture the structural breaks.

In addition to credit risk analysis, the Hidden Markov Model is also utilized in option pricing. Chuin Ching & Tak Kuen (2010) proposed a method to value European Call Option with Hidden Markov regime switching. The model is trained with the appreciation rate, interest rate and the volatility of to-be-priced risky asset; thus, the states in the Hidden Markov Chain are interpreted as the state of an Economy. Robert & Tak Kuen (2013) discusses the pricing of European Call Option in a pure-jump asset pricing model, in which the state of the economy is governed by the Hidden Markov Chain.

The Hidden Markov Model is also applied in topics of optimal investment model and asset allocations. The recent study by Shangzhen & Xudong (2014) extended the classical investment model by Merton, in which the volatilities of risky assets are assumed to be constant. Under new model proposed by Shangzhen &

Xudong (2014), both risky assets return and volatilities are assumed to be governed by Hidden Markov Chain.

2.2 Hidden Markov Model in Financial Time Series Forecasting

The usage of Hidden Markov Model in Financial time series can be traced to a decade ago. Hassan, R. (2005) proposed a Hidden Markov Model with continuous emission to forecast the next day stock closing price of 4 different airlines stock. The model he proposed used the intra-day high, low, open and closing price of stock to predict the next day closing price. However, the result was similar to performance of Artificial Neural Network and was unreliable in practical use. In order to improve the performance and accuracy of price prediction, Hassan, R. (2009) combine the Hidden Markov Based prediction method with fuzzy model to improve the accuracy of the model. The study applied the same data set from the previous research for both training and testing. The result showed an improvement in prediction error in comparison to original HMM based prediction model, Artificial Neural Network and ARIMA. In the latest iteration, Hassan, R. (Hassan R. , 2013) improved the system by introducing the Adaptive Fuzzy Interference System which allowed the system to be able to adapt to the new arrival of data. The author applied the new system with 5 consecutive weekly stock index price data vectors to predict the weekly index movement and the result showed improvement in accuracy over the previously proposed HMM-Fuzzy Model.

There were other researches proposed improvement or other approaches in training the Hidden Markov Model. Satish & Jerry (2010) compares the performance of prediction of Hidden Markov Model with Support Vector Machine in predicting the closing price of stocks in the S&P 500 Index. The paper proposed the k-mean algorithm for parameters initialization of Hidden Markov Model. The importance of initialization can be observed from the result; the hit rate of stock prediction decreased substantially. The initialization problem of Hidden Markov Model was not addressed in system proposed by Hassan, R. (2013), and thus further effort in investigating parameters initialization might be crucial to the prediction power of the model. Another research by Patrik, I. (2008) tried to build algorithmic trading strategy by applying both discrete and continuous Hidden Markov Model to predict the exchange rate of EURUSD. The author was able to make positive cumulative profit and obtain the Sharpe ratio 0.91 during the simulation period. Other than the exchange rate, the author also attempted to include other factors into the model. However, the result showed that the additional factors did not improve the model and worsen the profitability.

2.3 Other machine Learning Model applied in Financial Time Series Forecasting

Other than Hidden Markov Model, Other machine learning models were often applied in forecasting financial time series. In the previous studies on Thai Capital Market, Sittipong, S. (2012) attempted to predict the SET50 Index with Neuro-Fuzzy system, but was unable to achieve reliable prediction and concluded that technical indicators were not able to rapidly respond to the change of market direction. Thapanun, P. (Thapanun, 2013) attempt to forecast the stock market movement by using hybrid models, i.e. Support Vector Machine, Probabilistic Neural Network, and Back-propagation Neural Network with weekly closing price and Macroeconomic and technical factors. The author concluded that trading with hybrid model is able to outperform the returns of market indices in term of profitability.

2.4 Order Imbalances and Stock Return

Volume is often used as a proxy in literature to describe the relationship between trading activity and market return. However, the order imbalance bears more information in term of trader's intent and direction of the stock price is headed.

The empirical evidence from the earlier research on the relationship between individual return and order imbalances by Chordia T, and Subrahmanyam A (2002) based on the daily NYSE data indicates that traders tend to split orders over period to mitigate price impact, which causes autocorrelated price pressure and results in a predictable relation between the imbalance and equilibrium price changes. The later research by Chordia T. and Roll R. (2005) also revealed that the future stock return can be predicted by the lagged order imbalance over the interval from 5 to 60 minutes; this evidence supports that the market is not efficient in the strong form. The further research by Chordia T. and Roll R. (2008) on stock return, order flow and market liquidity reveals that the predictability of individual stock return tends to disappear when the market is in a more liquid regime.

CHAPTER 3
THEORETICAL FRAMEWORK

This chapter will briefly describe financial concept used in this study.

3.1 Efficient Market Hypothesis

The conventional investment theory proposed by Fama (1970) defined the efficient market as a market in which the prices always reflected the available information in three different considerations. In short summary, for weak form efficiency, the assets prices fully reflect the historical price; for the semi-strong form efficiency, the assets prices reflect all information that is publicly available; for the strong form efficiency, the assets prices reflect privileged information that is available to only specific participants. Consequentially, the result of such remark is that, in an efficient market, the prices of risky assets should accurately reflect the fundamental value, and thus no excess return can be generated from trading.

The empirical evidence over daily horizon seems to support the efficient market hypothesis; the previous literature by Chordia et al. (2005) shows that S&P Index follows random work and had insignificant auto-correlations despite the fact that public unavailable information was incorporated.

3.2 Market Efficiency, order imbalance and market liquidity

In an early research of market order imbalances on the S&P 500 by Chordia et al (2002) documented an interesting phenomenon; the market order imbalances (defined as daily aggregated purchase order less sell order) are highly predictable on the daily basis. Empirically, a day with high order imbalance will likely be followed by high order imbalance on the same side. However, given the predictability, the S&P 500 follows random work over a horizon of a day and had no auto-correlation at first or other longer lags. The observation implies that, some investors were able to correctly forecast the price pressure created by the order imbalances and exploit the price pressure, in which the trades are able to remove the auto-correlation of return within the horizon of one day.

Such phenomenon raises the question of how quickly the predictability of return is removed by the countervailing trades conducted by the investors who observed the order imbalance. However, it is certain that the process of removal of predictability of return is not instantaneous; it must take at least some time for investors to realize the information of order imbalances.

The further research by Chordia et al (2005) investigated the time taken for traders to take countervailing position that removes the predictability of returns. The result reconcile the belief that traders though do not have the information of order imbalance, but become aware of the information and take the countervailing position. Under the horizon of 30 minutes, the return is no longer predictable by using the order imbalances.

The empirical evidence also indicates that the speed of convergence is affected by the market liquidity. Chordia et al (2008) investigated the predictability of return using order imbalance in different liquidity regime. The result shows that the market in a more liquid regime is less predictable and is close to random walk. This observation implies in a liquid regime, information is more effectively incorporated into the price of risky assets. One rationale to explain this phenomenon is that due the smaller bid-ask spread in the liquid regime, informed traders have more incentive to submit the countervailing orders and thus catalyzed the speed of convergence.

CHAPTER 4
METHODOLOGY FRAMEWORK

4.1 Microstructure of SET and KOSPI

The trading system in both Thailand and Korea is known as the order driven market; in this system, the market operates without the intermediaries known as the market makers. Both sellers and buyers have to submit the prices and the quantity of securities they are willing to buy or sell to the brokerages. These orders are then submitted digitally from the brokerages to the order matching system of exchanges. The order matching of buy and sell orders is then done by using the order precedence rule, in which the orders are matched by using price as first priority and time as second priority.

4.1.1 Daily Price Limitation

To prevent price manipulation and protect investors from sudden price fluctuation, both SET and KRX impose an upper bound and lower bound to which the price of listed securities can move in a day. The KRX adopts the limitation of +15% and -15% bound calculated on each trading day; On the other hand, the SET allows prices of securities to fluctuate in the range of 30% of the previous daily closing price.

4.1.2 Tick Size

The tick size of stocks in difference price range can be summarized in the following table 4.1 and 4.2:

Table 4.1: Tick Size of SET listing

Price level (Baht)		**Tick Size (Baht)**
	Less than 2	0.01
2 or higher	Less than 5	0.02
5 or higher	Less than 10	0.05
10 or higher	Less than 25	0.10
25 or higher	Less than 100	0.25
100 or higher	Less than 200	0.50
200 or higher	Less than 400	1.00
400 or higher		2.00

Source: Stock Exchange of Thailand

Table 4.2: Tick Size of KRX listing

Price level (Wan)		Tick Size (Wan)
	Less than 1,000	1
1,000 or higher	Less than 5,000	5
5,000 or higher	Less than 10,000	10
10,000 or higher	Less than 50,000	50
50,000 or higher	Less than 100,000	100
100,000 or higher	Less than 500,000	500
500,000 or higher		1,000

Source: Korea Exchange

4.2 Order Imbalance

Based on the previous literature, the stock price movement can be predicted by the order imbalance indicator over a very short horizon. The research by Chordia et al (2005) defined the order imbalance in 3 different forms: the number of buy order less the number of sell order (OIB#), the number of buy-initiated shares purchased less the number of seller-initiated shares sold (OIBSh) and the dollars paid by buy-initiators less the dollars received by sell-initiators (OIB$). The last two factors OIBSh and OIB$ have empirically better predictability of future return in comparison to OIB#, but all three informations are only available to market makers or traders who are able to estimate the imbalance in the New York Stock Exchange.

For the target markets of this study, we have both sell order and buy order data widely available to the public, and hence the order imbalance indicator will be constructed based on the available information. We approximate our order imbalance indicator in a similar approach to the recent research by Shen D. (2015) known as the Order Imbalance Ratio (OIR).

The OIR measures the size of buy order in relative to the sum of number of buy orders and number of sell orders at a specific time point. Thus, the low value of order imbalance ratio implies that there is lack of demand or excess of supply on a particular asset; whereas, the high value of order imbalance ratio implies that there is excess demand or lack of supply on a particular asset.

The order imbalance will be expressed as a relative term. The reason behind using this method to construct the indicator is that we can quantize the indicator with ease and scale down the indicator.

$$OIR_t = \frac{V_t^B}{V_t^A + V_t^B}$$

$$Where\ V_t^B = volume\ \ of\ current\ best\ bid\ price$$

$$V_t^A = volume\ of\ curret\ best\ ask\ price$$

4.3 Data Quantization

For discrete case of the Hidden Markov Model, the discretization process needs to be conducted to convert both return and order imbalance indicator into representative symbols.

We classify the price movement into two categories; the price moves down and price remains the same or price moves up. On the other hand, there is no clear guideline on discretizing the order imbalance ratio denoted OIR; therefore, we separate the order imbalance into 3 groups, which are the groups with OIR in the 25 percentile, OIR above the 75 percentile and the OIR that is between 25 and 75 percentile. The 25 percentile and 75 percentile are approximated by averaging the 25 percentile and 75 percentile of each stock at each frequency during the pre-study period. Table 4.3 reports the detail of data quantization for the discrete Hidden Markov Model.

Table 4.3: Data Quantization for discrete the Hidden Markov Model

Symbol	Return interval	Order Imbalance Ratio
1	$0\% <$	$< 25\%\ percentile$
2	$0\% <$	$25\%\ percentile \leq OIR \leq 0.75\%\ percentile$
3	$0\% <$	$OIR > 75\%\ percentile$
4	$\geq 0\%$	$< 25\%\ percentile$
5	$\geq 0\%$	$25\%\ percentile \leq OIR \leq 0.75\%\ percentile$
6	$\geq 0\%$	$OIR > 75\%\ percentile$

Noted: For symbol 1, 2 and 3, the return interval can be interpreted as negative price movement, i.e. $\Delta P < 0$; whereas, the return interval of symbols 4, 5 and 6 can be interpreted as price movement that is not negative, i.e. $\Delta P \geq 0$.

Table 4.4: 25% and 75% percentile of Order Imbalance Ratio

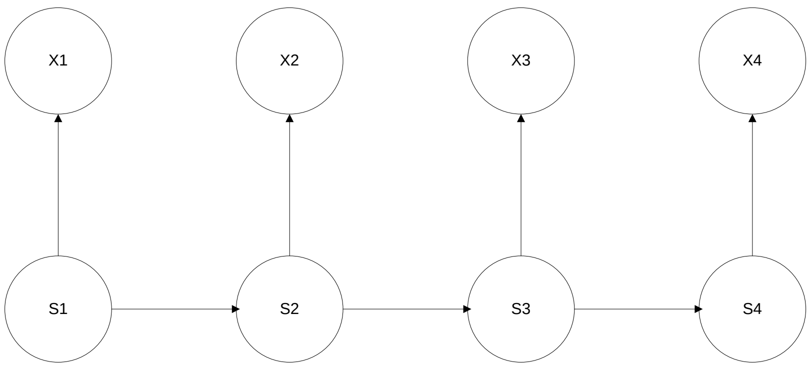
Market	Frequency	25% Percentile	75% Percentile
SET50	5 minute	0.40	0.65
	10 minute	0.40	0.64
	30 minute	0.41	0.62
KOSPI50	5 minute	0.34	0.61
	10 minute	0.35	0.60
	30 minute	0.38	0.59

Noted: The percentiles are computed based on the data from 1st October 2016 to 31st October 2016. The percentiles are computed for each individual stock then are averaged to obtain the value in the table.

4.4 Hidden Markov Model

The Hidden Markov Model (often referred as regime switching model or Markov Switching) is a statistical model that is designed to capture the dynamic that cannot be directly observed from a set of observations. The simple discrete Hidden Markov Model mainly consists of two parts, first a set of unobservable states $S = \{s_1, s_2, \ldots, s_T\}$ and a set of observable symbol $O = \{o_1, o_2, \ldots, o_T\}$. At each step/time slot t, the movement of state to another state is governed by a set of transition probability. The sequence of observable symbols is a state dependent process, i.e. each state governs a probability distribution of observable symbols. The following diagram demonstrates the states transition and observable symbols generated from the hidden states.

Figure 4.1: A graph of a simple Hidden Markov Model



The reason of applying Hidden Markov Model in this study is because of its ability to capture the hidden dynamic or behavior of stock market. In this study, we aim to capture the hidden state of order imbalances through the observable symbols of stock price movement and buy/sell order movement in a confident manner. The state of order imbalance can be interpreted as a state where new information has not yet adjusted into the asset price or the state where the asset price deviated from the fundamental. If the model is able to capture the order imbalance state in a consistent and confident manner, then it is possible to profit from the price pressure created by the order imbalance state.

4.4.1 Three Fundamental Problems of Hidden Markov Model

The characterization of a Hidden Markov Model can be described as following: 1) Number of states in the Model 2) Number of observable symbols in the model 3) the probabilities of state transition 4) the emission probability distribution of observable symbols generated from states 5) the prior probability distribution of initial states. For the rest of paper, following notations will be used.

N =	Number of states in the model
M =	Number of observable symbols
T =	Length of observable symbols sequence
H =	A set of possible states in the model, H = $\{h_1, h_2, \dots, h_N\}$
O =	The observable symbols sequence, $O = \{o_1, o_2, \dots, o_T\}$
S =	The states sequence, $S = \{s_1, s_2, \dots, s_T\}$
A =	The $N \times N$ state transition matrix
B =	The $N \times M$ observable symbols emission matrix
α_{ij}=	The probability of transition from state i to state j
$b_j(o_t) =$	The probability of generating observation t at state j
Π =	The $1 \times N$ vector of prior probability of each state
$\pi_i =$	Initial probability of starting in state i
$\boldsymbol{\lambda}$ =	The Hidden Markov Model, consisted of A, B and π. $\lambda = (A, B, \boldsymbol{\pi})$

The three fundamental problems of a Hidden Markov Model are of the following:

1. The Evaluation Problem: Given the model $\boldsymbol{\lambda}$, Compute the probability of the observed sequence of symbols i.e. compute $P(O|\boldsymbol{\lambda})$.
2. The Decoding Problem: Given both the model $\boldsymbol{\lambda}$ and the observed sequence of symbols, what is the most likely state sequence?
3. The Learning Problem: Given observation sequence and possible parameters of model i.e. $A, B\ and\ \Pi$, adjust the parameters to find the model that best explain the observed sequence, i.e. find $\boldsymbol{\lambda}$ that maximizes $P(O|\boldsymbol{\lambda})$.

The evaluation problem is used in the learning problem to test for convergence to the local maxima. The forward or backward algorithms are used to solve this problem and are explained below:

Forward algorithm:

Define probability of a subset sequence of forward variables being in the given state at time t

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, s_t = h_i|\lambda)$$

1. Initialization

   For $1 \leq i \leq N$

$$\alpha_1(i) = \pi_i b_i(o_1)$$

2. Iteration

   For $t = 2,3, \dots, T\ and\ 1 \leq j \leq N$

$$a_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i)\alpha_{ij}\right] b_j(o_t)$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^{N} a_T(i)$$

Backward algorithm:

Define probability of a subset sequence of backward variables and a given state at time t

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, s_t = h_i|\lambda)$$

1. Initialization

   For $1 \leq i \leq N$

$$\beta_T(i) = 1$$

2. Iteration

   For $t = T-1, T-2, \dots, 1$ and $1 \leq i \leq N$

$$\beta_t(j) = \sum_{i=1}^{N} a_{ij} b_i(o_{t+1}) \beta_{t+1}(i)$$

3. Termination

$$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i)$$

The decoding problem finds the most likely state sequence given the model and observation sequence. In this study, the Viterbi algorithm will be applied to solve the problem; it is an algorithm that finds the state sequence of a fixed observation sequence with the maximum likelihood i.e. $argmax\, P(S, O|\boldsymbol{\lambda})$**.** The Viterbi algorithm is defined below:

Define variable $\delta_t(i)$ as the maximum probability for sequences that end in states i and time t, i.e.:

$$\delta_t(i) = \max_{s1, s_2, \dots, s_{t-1}} P(s_1, s_2, \dots, s_t = h_i, o_1, o_2, \dots, o_t|\lambda)$$

By using induction, we have:

$$\delta_{t+1}(j) = \max_{i} \{\delta_{t-1}(i) \alpha_{ij}\}\, b_j(o_{t+1})$$

Having $\delta_t(i)$, we can determine the most probable state at time t:

$$s_t = \arg max\ \delta_t(i), for\ 1 \leq i \leq N$$

1. Initialization step:

$$\delta_1(i) = \pi_i b_i(o_1), i = 1,2, \dots, N$$
$$\theta_1(i) = 0$$

The array $\theta_t(i)$ is for keep tracking the $t-1$ state that maximizes $\delta_t(i)$

2. Iteration: $for\ 2 \leq t \leq T, 1 \leq i \leq N$

$$\delta_t(j) = \max_i\{\delta_{t-1}(i)\alpha_{ij}\}b_j(o_t)$$
$$\theta_t(j) = argmax\{\delta_{t-1}(i)\alpha_{ij}\}$$

3. Termination:

$$s_T^* = argmax\{\delta_T(i)\}$$
$$s_t^* = \theta_{t+1}(s_{t+1}^*), for\ t = T-1, T-2, \dots, 1$$

Last but not least, the learning problem of Hidden Markov Model finds the model parameters $\boldsymbol{\lambda}$ that best explain the observed sequence (maximizing the probability $P(O|\lambda)$). The learning problem cannot be solved analytically, and is conventionally solved by the Expectation Maximization algorithm called the Baum-Welch algorithm.

The Baum-Welch algorithm is an iterative process to approximate convergence to local optima. The Baum-Welch algorithm is explained in the following:

Define $\gamma_t(i)$ to be the probability of being in state i at time t

$$\gamma_t(i) = P(s_t = h_i|O, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)}$$

Where $\alpha_t(i) and\ \beta_t(i)$ are the forward and backward variables explained in the evaluating problem.

Define $\xi_t(i,j)$ to be probability of being in state i at time t and state j at time $t+1$

$$\xi_t(i,j) = P\big(s_t = h_i, s_{t+1} = h_j \big| O, \lambda\big) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{j+1}(j)}{P(O|\lambda)}$$

$$= \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{j+1}(j)}{\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{j+1}(j)}$$

The steps of algorithm are detailed below:

1. Initialization step:
   a. randomizing input parameters $\lambda = \{A, B, \Pi\}$,
   b. Set up a tolerance value tol
   c. A real number Δ
2. Iteration step:

   Iterate until $\Delta < tol$

   I. Calculate $P(O|\lambda)$ by using forward or backward algorithm

   II. Re-estimate λ^*: for $1 \le i \le N$

      a. $\pi_i^* = \gamma_1(i)$

      b. $a_{ij}^* = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}, for\ 1 \le j \le N$

      c. $b_i^*(k) = \frac{\sum_{t=1,o_t=k}^{T}\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)}, for\ 1 \le k \le M$

   III. Calculate $\Delta = |P(O|\lambda^*) - P(O|\lambda)|$

   IV. Update λ

If the observations are continuous and the probability distribution is Gaussian, the input parameter $\lambda = \{A, \mu, \sigma, \Pi\}$, the formulas for calculating $\mu\ and\ \sigma$ are as follow:

$$\mu_i^* = \frac{\sum_{t=1}^{T}\gamma_t(i)o_t}{\sum_{t=1}^{T}\gamma_t(i)}$$

$$\sigma_i^* = \frac{\sum_{t=1}^{T}\gamma_t(i)(o_t-\mu_i)(o_t-\mu_i)^T}{\sum_{t=1}^{T}\gamma_t(i)}$$

However, there is no guarantee of convergence and approximation of global optima. With the advance of computing power, sampling method known as gibbs sampling (a type of Markov Chain Monte Carlo) can be applied to achieve global optima of model. However due the computational complexity, we limit the scope to the Expectation Maximization ony.

4.4.2 Number of states in the Hidden Markov Model

As discussed in the literature review section, the number of states in the Hidden Markov Model can be interpreted as different behavior of markets. Determining the optimal number of states in the market would be crucial to the trading signal generation of the model. The number should not be too large, there is little to no distinction between each state; on the other hand, if the number of states also should not be too small, then the model may not be able to capture the hidden behaviors of market movement. For this study, we set the minimum number of states of stock to three, in which the three states represent the information of asset price being overvalued, undervalued or in the equilibrium. However, there are possibly unknown hidden states in the market; the model might be improved if we increase the number of states for coverage of other hidden states. For the scope of this study, we aim to test the performance of our model from 3 states to 5 states.

4.4.3 Generating Trading Signals

4.4.3.1 Discrete Case

From the discussion in the literature review section, we can use the solution to the learning problem to approximate the best model for the given observation sequence. By using the model, from the decoding problem we can find the probability of each state that generate the current observation and find the most probable state that generates the current symbol.

By knowing the most probable state at t, we can utilize the transition matrix A estimated in the learning problem to find out the likely transition and predict the state at $t + 1$. Then, based on the predicted state, we determine the probability of observing certain asset price movement by using the emission matrix B. In order to be more certain about the outcome in the next time period, a threshold needs to be imposed.

For the purpose of this study, we are interested in observing a state that has confident transition and follow by a state where the probability of upward price movement is high. Therefore, based on table 4.3, the threshold can be set as the probability of observing symbol 4, 5 and 6 at $t + 1$. The value of certainty of outcome at $t + 1$ can be determined as follow.

$$P\left(s_{t+1} = h_i \middle| s_t = h_j\right) \cdot b_i(o_t)$$

If the p-value of is greater than defined threshold, then the trading signal of entering position is generated, else liquidate current position. In this study, we want to capture the order imbalance state in a consistent and confident manner; thus, it would be in our interest to filter out the signals with lower confidence level to avoid excess loss from transaction cost and incorrect predictions. To set up our threshold, we propose a 90% confidence level in transition and 90% confidence level in observing positive or no price movement. The joining two confidence level, we propose to set the threshold at value of 80%.

4.4.3.2 Continuous Case

For the continuous case of Hidden Markov Model, each hidden state is associated with the probability density function of observables instead of discretized probability for each possible observable. Therefore, the trading signals are generated based on the interpretation on the properties of probability distribution function.

We report the summary statistic and normality test (Sharpiro-Wilk Normality test) in appendix A and B, the result indicates that the probability distribution function of return and OIR are not normal with 95% confidence level. We also examined the multivariate normality of {return, order imbalance ratio} by using the Henze-Zirkler Multivariate Normality test and reported the result in appendix B; and similarly, we reject the null hypothesis that the probability distribution is normal for every stock at every intra-day frequency with 95% confidence level.

With support of empirical evidence that the probability distributions of observables are not normal, we move our attention to the Gaussian Mixture Model to better describe the properties of probability distribution of observables.

4.4.3.2.1 Gaussian Mixture Model

The Gaussian Mixture Model (GMM) treats observations as if they are coming from various sources and each source is modeled as Gaussian distribution. Thus, each source has its own set of parameters (μ_i and Σ_i) of probability density function, and the weighted sum of each source is the Gaussian Mixture Mode. Formally, the Gaussian Mixture Model is a probability density function built up from weighted sum of M Gaussian components (see figure 4.2 for an example of Gaussian Mixture Model) and is defined by the following equation:

$$P(X|\theta) = \sum_{i=1}^{M} w_i G(X|\mu_i, \Sigma_i),$$

Where X is a D-dimensional vector of data,

$w_i, i = 1,2, \dots, M$ are the weight of each Gaussian component in the mixture.

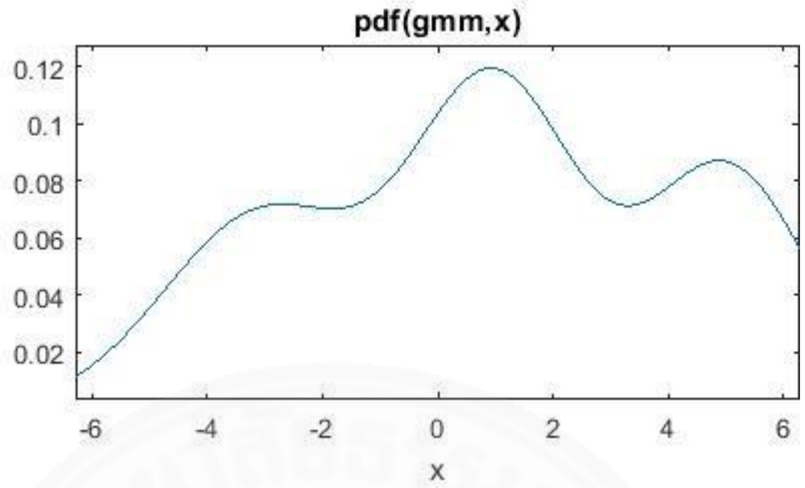
$G(X|\mu_i, \Sigma_i)$ is the D-variate Gaussian component with form:

$$G(X|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_i)' \Sigma_i^{-1} (X-\mu_i)}$$

Where μ_i is the mean vector of the of multivariate Gaussian component

Σ_i is the covariance matrix of multivariate Gaussian component

Figure 4.2: Example of Gaussian Mixture Model of 3 Components



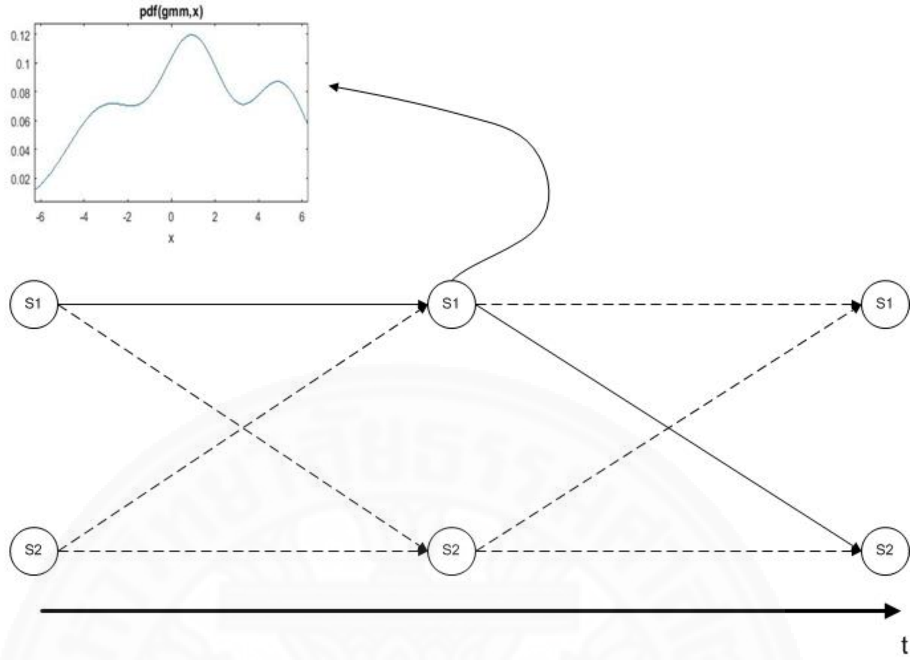
To determine the number of Gaussian components that is suitable for this study, we performed the iterative Expectation Maximization algorithm on datasets of {return, OIR} on each stock at different frequency levels. We report the result in appendix C and the result shows that for majority of stocks at different frequency, the 3-components Gaussian Mixture Model has the lowest BIC score and thus is chosen as the probability density function for the hidden states.

4.4.3.2.2 HMM with the Gaussian Mixtures

As discussed in the previous sub sections, the observables are continuous. Therefore, instead of the emission matrix of observables, we have the parameters for the probability density function of Gaussian Mixture Model as shown in figure 4.3. As a result, the probability of observables generated from a particular state at time t defines as:

$$b_j(o_t) = \sum_{i=1}^{M} w_i P(o_t | G_j\ (O|\mu_i, \Sigma_i))$$

Figure 4.3: 2 states HMM with Gaussian Mixture Model of 3 components



4.4.3.2.3 Generating Trading Signal

In this study, we propose 2 approaches to generate trading signal and are discussed below:

Approach I: Using only first moment:

To generate trading signal, we first set-up a threshold level of return 0**.** At each trading interval, the Viterbi algorithm is first used to determine the most probable state at current t. Then, we utilized the trained transition matrix to determine the state at $t+1$. The expected return is then calculated by using the mean return of each Gaussian component. The equation is defined as follow:

$$E[r] = \sum_{i=1}^{M} w_i \mu_{i,return}$$

If the expected return is greater than 0, then trading signal is generated; whereas, if the expected return is less than 0, then we liquidate the position.

Approach II: Using both first moment and second moment

To generate the trading signal, we first set-up 2 thresholds, a threshold of confidence level of next period return 0.8 and a threshold for expected return 0**.** At each trading interval, the Viterbi algorithm is firstly used to determine the most probable state at current time t. Then, we utilized the trained transition matrix to determine the future state at $t + 1$. The p-value is then calculated by using the following equation:

$$p - value = P\left(s_{t+1} = h_i \middle| s_t = h_j\right) \cdot \sum_{i=1}^{M} w_i \, P(return \geq 0 | G(X|\mu_i, \Sigma_i))$$

If calculated p-value is greater than or equal to the defined threshold, then the trading signal is then generated.

4.5 Trading Strategies

In the previous section, the paper has discussed about how the signal is generated from the Hidden Markov Model. Due to the nature of discretization method, the generated signal can only predict the direction of movement, but not the size movement. Thus the strategy is a form of gambling with the belief that there will be more gains than losses from the gambling. In this study, we propose a simple algorithm to handle the signals generated from the Hidden Markov Models.

Strategy I: Equally Weighted Allocation

1. Train the Hidden Markov Model for each stock
2. Obtain a list of stocks (trading signals) that we should enter long position.
3. Liquidate all stocks that are current in long position and are not in the list.
4. If there is any remaining wealth, allocate the wealth equally to all stocks in the list.
5. If at the end of the day, then go to step 1, else go to step 2
6. Iterate until the end of observations

The mid-point closing price at the end of each interval will be used as the trading price for buying and selling the shares and bi-directional transaction cost at level of 0%, 0.05% and 0.1% are used to assess applicability of the model for different group of investors.

4.6 Performance Measurement

4.6.1 Benchmark

The SET and KOSPI Total Return Index will be used as benchmark to compare with the profitability of the trading strategy. The Total returns index is calculated based on the assumption that all dividends are immediately re-invested.

4.6.2 Hit Ratio

The hit ratio will be used to assess the performance of Hidden Markov Model on forecasting the stock price movement of out-of-sample data set. The calculation of Hit Ratio will be separated from the trading strategy and will be calculated for each individual stocks.

The hit ratio is defined as follow:

$$Hit\ Ratio = \frac{h}{n}$$

Where n = total number of trading signals that results in positive/negative price movement

h = total number of trading signals that correctly predict the positive price movement

To test whether or not the forecast is more than just coin flip guess, we will apply one sample t-test to test the null hypothesis whether or not the hit ratio is equal to 0.5. The formula of one sample t-test is defined as follow:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Where $\bar{x}$ = sample mean

$\frac{s}{\sqrt{n}}$ = standard error

We will use one-sample t-test to test the hypothesis of $H_0: Hit\ Ratio = 0.5, H_a: Hit\ Ratio \neq 0.5$. Under Efficient Market Hypothesis, the null hypothesis should not be rejected; if the null hypothesis is rejected, then there exists pattern in the stock market and the price movement can be predicted.

4.6.3 Sharpe Ratio

In order to compare with the benchmark buy-and-hold strategy, the Sharpe ratio will be used. The Sharpe ratio measures the risk premium over the amount of risk, higher the Sharpe ratio, the more desirable the asset is. The ratio is calculated as follow:

$$Sharpe\ Ratio = \frac{E(r_i) - r_f}{\sigma_i},$$

Where $E(r_i)$ = Expected return of risky asset

r_f = risk free rate

σ_i = volatility of the risky asset

To fit the horizon of our study, the 3-Month BIBOR (Bangkok Interbank Offered Rate) and 3-Month KORIBOR (Korea Interbank Offered Rate) will be used as our proxy to risk-free rate in the computation of Sharpe Ratio.

The Sharpe Ratio will be calculated based on the monthly return of both trading strategies and the SET and KOSPI Total Return Index.

4.6.4 Jensen's Alpha

Alpha is the abnormal rate of return that exceeds the expected return at given risk defined by a specific model. Under the efficient market hypothesis, the alpha should be insignificant and equal to 0, because it is not possible to outperform the overall market. For this study, the Capital Asset Pricing Model (CAPM) will be used to estimate whether there exist significant alpha for our back-testing portfolio over the horizon of three months. The model is defined as:

$$E[r_i] - r_f = \alpha_J + \beta(E[r_M] - R_f)$$

Where r_i = portfolio return

r_f = risk free rate

r_M = market return

4.7 Data

We conduct this study in markets with different level of liquidity: the Thai stock market and Korean stock market. In particular, the Korean market is more liquid than the Thai capital market. As reported by the World Bank, in year 2015, the annualized stock turnover ratio of Stock exchange of Thailand (SET) is roughly 77.8%; in contrast, the annualized stock turnover ratio of Korea Exchange (KRX) is roughly 149.8%, which is more than 2 times of turnover ratio of Thai stock market. Based on the theory, we expect that the Korea Exchange should have lower predictability of asset return due having higher liquidity; in other word, our model should perform relatively poor in Korea stock market in comparison to Thai stock market.

We limit the scope of this study to stocks listed in SET50 Index and KOSPI50 Index; the SET50 Index is chosen because the stocks are relatively more liquid in comparison to other stocks listed in the SET and are the stocks with large market capitalization, thus mitigate the issues of no trades. To compare between markets, we decide to pick KOSPI50 index in Korea that selects the stocks in similar method of SET50. To further limit the scope of this study, we reduce the number of stocks to 10 for each market and the selection method is described below.

The stocks in this study are selected by going through the following steps

1. Filtered stocks that are not consistently listed in SET50 Index and KOSPI 50 Index during the period form 1st January 2012 to 31st July 2016
2. Keep the top 10 stocks with highest average volume turnover in the respective market to ensure liquidity of stocks. The turnover is calculated by using the following formula**:**

$$Volume\ Turnover = \frac{250\ days\ average\ daily\ volume\ turnover\ (as\ of\ 21st\ November\ 2016)}{Average\ Total\ Common\ shares\ outstanding(Through\ out\ year\ 2015)}$$

The list of stocks applied in this study is represented in the table 4.5 and 4.6.

Table 4.5: The listed stocks selected from SET50 for study

Ticker	Company Name	Sector
ADVANC.BK	Advance Info Service PCL	Information & Communication
BANPU.BK	Banpu PCL	Energy & Utilities
BCP.BK	Bangchak Petroleum PCL	Energy & Utilities
CPF.BK	Charoen Pokphand Foods PCL	Food and Beverage
DTAC.BK	Total Access Communication PCL	Information & Communication
IRPC.BK	IRPC PCL	Energy & Utilities
IVL.BK	Indorama Ventures PCL	Petrochemicals & Chemicals
PTTEP.BK	PTT Exploration and Production PCL	Energy & Utilities
TCAP.BK	Thanachart Capital PCL	Banking
TRUE.BK	True Corporation PCL	Information & Communication

Table 4.6: The listed stocks selected from KOSPI 50 for study

TICKER	Company Name	Sector
034220.KS	LG Display Co, Ltd	Electrical & Electronic Equipment
066570.KS	LG Electronics Inc	Electrical & Electronic Equipment
051910.KS	LG Chem Co, Ltd	Chemicals
005490.KS	POSCO	Iron & Metal Products
006400.KS	Samsung SDI Co, Ltd	Electrical & Electronic Equipment
009150.KS	Samsung Electro Mechanics Co Ltd	Electrical & Electronic Equipment
010140.KS	Samsung Heavy Industry Co, Ltd	Transport Equipment
000880.KS	Hanwha Corp	Finance
000720.KS	Hyundai Engineering & Construction Co Ltd	Construction
009540.KS	Hyundai Heavy Industry Co, Ltd	Transport Equipment

As discussed in the previous section, the input of the model are bid size, ask size, closing price at interval of 5, 10 and 30 minutes. Based on these data, order imbalance indicator and return are computed. The model uses rolling window technique and the window will be move by 1 day for every step; the size of window are 15 trading days for 5 minute data, 19 trading days for 10 minute and 40 trading days for 30 minutes data. The training is conducted on the daily basis; this means that the model will be re-trained when the market closed. The actual size of rolling window is defined by:

$$Size = Number\ of\ interval\ per\ day \times number\ of\ days$$

The data will be divided into 2 periods: the pre-study period and the simulation period. The data in the pre-study period will be used for both initial training of model and estimation of 25 and 75 percentile of order imbalance ratio. The trading period will be the period for test both profitability and accuracy in this study. The pre-study period starts from 1st October 2016 to 31st October 2016; the study period begins from 1st November 2016 and ends in 31st January 2017.

For performance comparison, we collected the daily and monthly 3-month Bangkok Interbank Offered Rate (BIBOR) and Korea Interbank Offered Rate (KORIBOR) as the proxy to risk-free rate. The daily and monthly total return index of Stock Exchange of Thailand and Korea Exchange (KRX) are collected as our benchmark.

The intra-day data of selected stocks are collected from Reuter Eikon, the daily and monthly 3-month BIBOR rate is collected from database of Bank of Thailand. The daily and monthly 3-month KORIBOR rate is collected from Korean Statistical Information Service.

CHAPTER 5
RESULT AND DISCUSSION

In this chapter, we report the result in term of predictability of the model and the profitability of strategy. We first report the observed predictability of models for both continuous and discrete Hidden Markov Models with different frequency and in different markets in section 5.1 and 5.2, and then discuss the possible causes of different result in term of different models, frequency and level of liquidity. In section 5.3, we discuss the profitability of the strategy at different level of transaction cost.

5.1 Predictability of the discrete model

The discrete models are proven to be capable finding the order imbalance state of selected stocks in Thai and Korean market, but with different performance. The result of our models seems to be consistent with the literature that the convergence to market efficiency is not an instantaneous process, but rather, a process that takes a short period of time; and market liquidity should enhance the speed of convergence.

5.1.1 Predictability of the Discrete HMM for selected stocks in SET50

Based on table 5.1, the models perform more confidently and consistently in the Thai market; our basic case, the basic 3 states discrete Hidden Markov Model is able to achieve hit ratio of average 78.61% at 5 minute frequency. The model seems to improve at 5 minute frequency as we increase the number of states; at 5 states, the average hit ratio increases to 83.38% with no predictability lower than 70% for each individual risky asset.

As we lower the frequency, the hit ratio decreases and the models become less confident in making a prediction. Moving from 5 minute frequency to 10 minute frequency, the average hit ratio lowers to 70.59%. The total number of predicted signals also decreases by approximately a factor of 5 (see appendix F). At frequency of 30 minutes, the average hit ratio of discrete models decreases to 65.63%.

The result seems to support the hypothesis that the order imbalance tends to lose its predictability as the interval enlarge; our model becomes less confident in making predictions (in the form of generating less signals) with cases that no prediction was made. This shows that, as the time increases, it becomes more difficult for the Hidden Markov Model to recognize a pattern. This evidence is consistent with the previous literature by Chordia et al (2005) that the information is adjusted into the price as time increases, thus the individual stock price becomes less predictable and follows random walk.

5.1.2 Predictability of the Discrete HMM for selected stocks in KOSPI50

In table 5.2, we report the result of our models in the Korean stock market produces evidence that the market liquidity does enhance the speed of convergence to efficiency. Compared to the model performance in Thai market, the models though were able to achieve some predictability, seem to be much less confident in making predictions. Begin with the 5 minute frequency; the models generate relatively low number of signals in comparison to our model performance in the Thai market. Though the daily trading period in the Korean capital market is longer than the Thai market, the total number of signals generated is significantly less; over the horizon of 3 month, the average total number of signals generated for the Korean market is 289 signals (see table F.6 in appendix F). In addition, the 3 states and 4 states model were only able to generate signals for less than half of the selected stocks. At interval of 30 minutes, the average total number of signals decreased to 71 signals over the horizon of 3 months.

In term of hit ratio, the result indicates better models performance in lower frequency and tends to probability of coin toss in lower frequency. At 5 minute frequency, the average hit ratio of 3 states, 4 states and 5 states models is approximately 67.74% (see table 5.4), with the 5 states model performs in a more consistent manner (generated the most signals and achieved average hit ratio of 71.57%) . As the frequency decreases, the hit ratio indicates that the signals generated by models were no longer able to predict the price movement.

The result is consistent with the previous literature that the market liquidity enhances the speed of convergence to efficiency. Our models are significantly less confident in generating a signal and achieve low hit ratio in comparison to the same models performance in the Thai market. The finding seems to be consistent with the literature by Chordia et al (2008); when the market is in a more liquid regime, the bid-ask spread tends to narrower; as a result, the traders who observe the order imbalance will then have more incentives to take position and gain from the deviation of asset price from the fundamental. Such actions enhance the speed of adjustment of asset price, and our models become less confident in capturing the price pressure created by order imbalance due to the fact that the process of price adjustment to new information already occurred.

Table 5.1: Hit ratio of the Discrete HMM for selected stocks in SET50

The table shows the hit ratio of the Discrete Hidden Markov Model on each risky asset selected from SET50 Index with different number of states at different level of frequency. The hit ratio is calculated by number of generated signals that correctly predicted the future positive movement divided by number of signals generated that result in either negative or positive price movement. Any signals that resulted in zero mid-point price movement are ignored. See table F.1 and table F.2 of appendix F for number of signals generated and number of signals that correctly predict the market movement,

Frequency	**# of states**	**ADVANC.BK**	**BANPU.BK**	**BCP.BK**	**CPF.BK**	**DTAC.BK**	**IRPC.BK**	**IVL.BK**	**PTTEP.BK**	**TCAP.BK**	**TRUE.BK**
5 min	3	82.47%	66.92%	81.19%	64.88%	82.72%	78.64%	78.17%	74.10%	89.68%	87.33%
	4	77.43%	75.00%	89.86%	80.30%	83.85%	83.18%	74.38%	71.67%	86.47%	83.82%
	5	80.12%	80.70%	89.42%	85.31%	84.77%	89.19%	80.59%	70.73%	88.46%	84.46%
10 min	3	72.22%	73.68%	91.55%	85.37%	57.58%	72.86%	70.27%	54.17%	71.43%	72.97%
	4	86.67%	53.85%	66.67%	69.05%	66.67%	100.00%	63.64%	-	-	72.73%
	5	73.21%	44.68%	85.71%	77.22%	73.44%	85.29%	62.26%	40.00%	66.67%	62.96%
30 min	3	60.00%	67.86%	87.50%	69.44%	72.22%	72.97%	78.43%	50.00%	83.87%	75.76%
	4	40.00%	25.00%	100.00%	40.00%	80.00%	57.14%	65.52%	75.00%	-	83.33%
	5	70.00%	53.85%	60.00%	50.00%	83.33%	62.50%	68.75%	71.43%	44.44%	57.69%

Table 5.2: Hit ratio of the Discrete HMM for selected stocks in KOSPI50

The table shows the hit ratio of the Discrete Hidden Markov Model on each risky asset selected from KOSPI50 Index with different number of states at different level of frequency. The hit ratio is calculated by number of generated signals that correctly predicted the future positive movement divided by number of signals generated that result in either negative or positive price movement. Any signals that resulted in zero mid-point price movement are ignored. See table F.5 and table F.6 of appendix F for number of signals generated and number of signals that correctly predict the market movement.

Frequency	**# of states**	**034220**	**066570**	**051910**	**005490**	**006400**	**009150**	**010140**	**000880**	**000720**	**009540**
5 min	3	-	-	78.13%	72.62%	67.39%	-	-	-	-	-
	4	-	-	66.67%	80.00%	0.00%	52.63%	-	-	-	95.35%
	5	53.97%	76.92%	80.56%	77.14%	72.73%	61.73%	66.67%	-	75.00%	79.44%
10 min	3	58.33%	42.42%	70.69%	59.52%	75.00%	-	100.00%	50.00%	-	73.33%
	4	20.00%	62.07%	57.30%	57.14%	80.00%	100.00%	100.00%	50.00%	-	100.00%
	5	83.33%	50.00%	55.74%	40.00%	50.00%	100.00%	40.00%	-	42.86%	72.73%
30 min	3	-	-	-	60.00%	-	-	-	-	-	-
	4	-	0.00%	71.43%	48.72%	83.33%	50.00%	-	-	0.00%	47.83%
	5	-	25.00%	66.67%	50.00%	66.67%	83.33%	33.33%	42.86%	0.00%	37.50%

5.2 Predictability of the continuous model

As discussed in section 4.4.3.2, we incorporate two approaches to generate signals for predicting the market movement. For the first approach, the signal is generated by using the mean return of the predicted state; whereas, for the second approach, the signals are generated by calculating the probability of observing positive price movement. All in all, regardless which approach we use, our result of continuous model indicates that the models failed to capture the order imbalance states across all frequencies and number of states.

The first approach, i.e. predicting the movement by the mean return of state, also failed to generate a meaningful result. As shown in table 5.3 and 5.4, the hit ratios of the continuous model for all stocks, across all frequencies, are around the number of 0.5. Unsurprisingly, our t-test result indicates that the hit ratios for all cases of continuous models are not statistically deviated from 0.5, and we failed to reject the null hypothesis that the hit ratios are equal to 0.5. This result indicates that the continuous Hidden Markov Model failed to capture the order imbalance state, and the predictability is the same as a coin toss.

On the other hand, the approach II i.e. making prediction based on the probability distribution of return, the models were unable to generate p-value higher than 60% in both markets; therefore, the model was unable to generate a single signal due to our requirement of capturing the order imbalance state in a confident and consistent manner and the result was not recorded.

One possible explanation of why continuous models failed to produce meaningful result is the assumption of distribution. Due to the fact that intra-day return and order imbalance indicator are not normal (see appendix B), we attempt to mitigate the issue by assuming the three components Gaussian Mixture Models. However, the resulting Gaussian Mixture Model might not be enough to mitigate the extreme Kurtosis value of intra-day data (see appendix C). Comparatively, the BIC score of the 3 components Gaussian Mixture Model is better than the Gaussian Model, but not significantly better. As a result of excess kurtosis, the models suffer from the assumption and thus failed to produce meaningful result.

Table 5.3: Hit ratio of the Continuous HMM for selected stocks in SET50

The table shows the hit ratio of the Continuous Hidden Markov Model on each risky asset selected from SET50 Index with different number of states at different level of frequency. The prediction method is the mean return of the state predicted by the Viterbi algorithm. The hit ratio is calculated by number of generated signals that correctly predicted the future positive movement divided by number of signals generated that result in either negative or positive price movement. For number of signals generated and number of signals that correctly predict the market movement, please see table F.3 and table F.4 of appendix F.

Frequency	**# of states**	**ADVANC.BK**	**BANPU.BK**	**BCP.BK**	**CPF.BK**	**DTAC.BK**	**IRPC.BK**	**IVL.BK**	**PTTEP.BK**	**TCAP.BK**	**TRUE.BK**
5 min	3	50.46%	49.04%	49.88%	49.57%	50.81%	50.35%	50.68%	50.08%	50.87%	48.85%
	4	50.90%	48.14%	50.11%	49.15%	50.00%	50.00%	51.23%	50.19%	50.80%	48.53%
	5	50.55%	48.29%	49.83%	49.59%	50.12%	50.00%	50.56%	50.87%	50.70%	48.28%
10 min	3	51.68%	47.65%	50.36%	47.58%	49.50%	50.00%	50.80%	48.63%	50.11%	48.98%
	4	51.07%	47.88%	49.85%	49.70%	48.97%	49.46%	49.68%	49.22%	49.44%	48.03%
	5	50.41%	47.60%	50.34%	49.83%	48.52%	49.89%	49.90%	48.71%	50.63%	48.99%
30 min	3	51.24%	47.46%	50.48%	42.42%	53.16%	52.38%	51.48%	48.68%	52.28%	46.67%
	4	51.75%	48.15%	49.65%	48.67%	52.82%	51.00%	50.15%	49.11%	53.31%	46.34%
	5	49.57%	46.54%	50.58%	44.88%	52.50%	50.97%	51.54%	49.34%	52.69%	47.85%

Table 5.4: Hit ratio of the Continuous HMM for selected stocks in KOSPI50

The table shows the hit ratio of the Continuous Hidden Markov Model on each risky asset selected from KOSPI50 Index with different number of states at different level of frequency. The prediction method is the mean return of the state predicted by the Viterbi algorithm. The hit ratio is calculated by number of generated signals that correctly predicted the future positive movement divided by number of signals generated that result in either negative or positive price movement. For number of signals generated and number of signals that correctly predict the market movement, please see table F.7 and table F.8 of appendix F.

Frequency	**# of states**	**034220**	**066570**	**051910**	**005490**	**006400**	**009150**	**010140**	**000880**	**000720**	**009540**
5 min	3	49.80%	51.42%	51.69%	50.13%	50.87%	49.79%	47.62%	48.80%	50.30%	49.17%
	4	50.24%	51.66%	50.59%	49.82%	50.86%	49.51%	48.56%	48.93%	51.55%	49.32%
	5	50.23%	51.63%	50.70%	50.10%	51.00%	49.48%	50.47%	48.55%	49.51%	48.74%
10 min	3	49.66%	51.43%	50.11%	49.22%	51.39%	50.11%	49.26%	48.54%	49.68%	49.13%
	4	49.70%	52.06%	50.66%	49.53%	51.47%	49.94%	48.24%	47.93%	47.46%	49.10%
	5	49.06%	51.98%	50.78%	49.62%	51.58%	47.41%	48.92%	49.39%	48.35%	49.05%
30 min	3	44.02%	53.54%	50.11%	48.80%	54.63%	48.08%	45.87%	49.63%	49.85%	47.95%
	4	46.86%	53.37%	49.88%	47.73%	54.05%	50.56%	47.74%	46.68%	49.50%	45.51%
	5	45.82%	53.87%	48.32%	47.98%	54.23%	50.14%	46.09%	46.63%	48.19%	45.33%

5.3 Performance of trading strategy

The second objective of this study is to create an algorithmic trading strategy based on the signals generated from the Hidden Markov Model. As discussed in chapter 4, the two matrices that we have used to measure the performance are the Jenson's Alpha and Sharpe's Ratio. For Jenson's Alpha, we computed the daily excess return of the strategy and the market, and then we conduct regression analysis to test for the significance of abnormal return. Whereas, we computed the Sharpe's Ratios of the strategy and the market based on the monthly volatility and mean return, and then compared which strategy achieves higher Sharpe's Ratio.

5.3.1 Profitability of the discrete models in both Thai and Korean stock market

As reported in table 5.12, with the assumption of 0.05% bi-directional cost, the result indicates that even at the highest frequency, our trading strategy was not able to achieve significant alpha in the Korean market. In contrast, as reported in table 5.11, the strategy shows a promising result of yielding significant positive alpha in all 5 minute cases and several cases at the 10 minute and 30 minute frequencies. The result shows that it is possible for institutional traders to make a profit by trading based on the predicted market movement. We also look at other scenarios with different assumptions on transaction cost.

Table E.1 and E.7 in appendix E report the Jenson's Alpha of the strategy with assumption of no transaction cost in Thai and Korean stock market; the trading strategy was able to beat the market at 5 minute frequencies in both markets. In both markets, the result of regression analysis indicates that our portfolio yields significant alpha up to 30 minute frequency in Thai market and 10 minute frequency in Korean market.

We report the Jenson's Alpha and Sharpe ratios of trading strategy with assumption of 0.1% bi-directional transaction cost in table E.3 and E.9 of appendix E and table D.3 and D.9 of appendix D. The profitability of the strategy in both markets almost disappeared with only the 5 states model at 5 minute frequency in the Thai market was still able to outperform the market with alpha of 1.45% on the daily basis.

This result indicates that this trading strategy is not suitable for non-institutional traders due to the profit being offset by the transaction cost.

However, it is to be noted that the result is under assumption of trading with mid-point price than the actual bid-ask price. Based on the previous literature, the reason of the strategy works in market with lower liquidity is because of lack of incentives for sophisticated traders to take position due to higher bid-ask spread.

5.3.2 Profitability of the continuous models in both Thai and Korean stock market

As discussed in section 5.1, the continuous models failed to capture the order imbalance state. Similarly, trading strategy using the continuous models failed to generate a significant and meaningful result.

We report Sharpe's Ratio of the strategy with institutional level of transaction cost in table 5.7 and 5.8; the Sharpe ratio of the trading strategy using continuous models can be either lower or higher than the benchmark, but there is no clear pattern on what is driving the portfolio return. Unsurprisingly, the results of our regression analysis provided in table 5.11 and 5.12 indicate that there exists no significant alpha for all models, across all frequencies and in both market.

To further evaluate the profitability of continuous model, we report the Sharpe's Ratio and Jenson's Alpha of the strategies with no transaction cost in appendix D and E. The result indicates that even at no transaction cost, there is no clear pattern to explain different value of Sharpe's Ratio across different frequencies and different models; expectedly, we observe no significant alpha from the strategy.

Consequentially, as discussed in section 5.1.2, the continuous Hidden Markov Models fail to capture to order imbalance states because our assumption on distribution is unable to describe the properties of intra-day data. As a consequence, our models are lack of predictability and are unable to obtain a significant return from the strategy.

Table 5.5: Sharpe ratio of trading of the Discrete HMM in SET50

The following table compares the Sharpe ratio between the discrete Hidden Markov Models with different number of states at length of interval. The Sharpe ratio is computed by the excess monthly mean return divide by the volatility of portfolio over the horizon from 1st November 2016 to 31st January 2017. As discussed in the methodology section, the benchmark for Thai market is the market return of SET and the adjusted monthly 3-month Bangkok Interbank Offered Rate is used for calculation of excess return. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table D.1 and D.3 in appendix D.

	5 minutes			10 minutes			30 minutes			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	62.4152%	76.5519%	99.0951%	23.3392%	3.2343%	10.7075%	16.2559%	0.8825%	1.7823%	2.0124%
σ	26.4160%	15.8375%	17.9793%	4.8111%	8.5142%	22.0272%	5.6533%	3.6992%	3.5969%	1.1801%
Sharpe Ratio	2.3590	4.8273	5.5061	4.8305	0.3682	0.4816	2.8579	0.2118	0.4679	1.6213

Table 5.6: Sharpe ratio of trading of the Discrete HMM in KOSPI50

The following table compares the Sharpe ratio between the discrete Hidden Markov Models with different number of states at length of interval. The prediction method of mean return (please see method I of section 4.4.3.2) is used for generating signal. The Sharpe ratio is computed by the excess monthly mean return divide by the volatility of portfolio over the horizon from 1st November 2016 to 31st January 2017. As discussed in the methodology section, the benchmark is the market return of KOSPI and the adjusted monthly 3-month Korea Interbank Offered Rate is used for calculation of excess return. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table D.7 and D.9 in appendix D.

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	-1.5605%	1.2178%	1.0058%	-2.1209%	-5.6903%	-3.4060%	-0.9012%	-1.7613%	-3.4782%	1.4687%
σ	3.3043%	1.0524%	8.6708%	2.8765%	4.0859%	2.7977%	1.5609%	1.0726%	3.1533%	2.1615%
Sharpe Ratio	-0.5098	1.0394	0.1017	-0.7804	-1.4230	-1.2617	-0.6567	-1.7575	-1.1423	0.6221

Table 5.7: Sharpe ratio of trading of the Continuous HMM in SET50

The following table compares the Sharpe ratio between the continuous Hidden Markov Models with different number of states at length of interval. The prediction method of mean return (please see method I of section 4.4.4.3) is used for generating signal. The Sharpe ratio is computed by the excess monthly mean return divide by the volatility of portfolio over the horizon from 1st November 2016 to 31st January 2017. As discussed in the methodology section, the benchmark is the market return of SET and the adjusted monthly 3-month Bangkok Interbank Offered Rate is used for calculation of excess return. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table D.4 and D.6 in appendix D.

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	0.1361%	-0.0948%	7.3997%	3.0263%	3.4718%	-0.1371%	3.2538%	1.1178%	4.8025%	2.0124%
σ	3.6048%	2.8341%	1.6411%	5.8075%	5.3254%	3.7583%	3.9488%	1.9440%	3.6249%	1.1801%
Sharpe Ratio	0.0102	-0.0684	4.4485	0.5040	0.6333	-0.0629	0.7989	0.5240	1.2975	1.6213

Table 5.8: Sharpe ratio of trading of the Continuous HMM in KOSPI50

The following table compares the Sharpe ratio between the continuous Hidden Markov Models with different number of states at length of interval. The prediction method of mean return (please see method I of section 4.4.4.3) is used for generating signal. The Sharpe ratio is computed by the excess monthly mean return divide by the volatility of portfolio over the horizon from 1st November 2016 to 31st January 2017. As discussed in the methodology section, the benchmark is the market return of the KOSPI and the adjusted monthly 3-month Korea Interbank Offered Rate is used for calculation of excess return. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table D.10 and D.12 in appendix D.

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	6.7943%	1.1443%	1.6548%	-0.0611%	0.4325%	2.6582%	2.5741%	1.4479%	3.5543%	1.4687%
σ	6.2202%	6.2382%	1.8425%	9.1623%	5.8634%	6.4040%	10.8685%	11.8982%	11.6520%	2.1615%
Sharpe Ratio	1.0763	0.1675	0.8443	-0.0175	0.0568	0.3996	0.2277	0.1134	0.2965	0.6221

Table 5.9: Jenson's Alpha (Discrete HMM, SET50, 0.05% transaction cost)

The following tables compare the trading performance of the discrete Hidden Markov Model in SET50 over the horizon of 1st November 2016 to 31st January 2017. The independent variable is the excess daily return of portfolio and the dependent variable is the excess daily market return of SET. The adjusted daily 3-month Bangkok Interbank Offered Rate is used as a proxy to risk-free rate. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table E.1 and E.3 in appendix E. All standard errors are Heteroskedasticity-robust standard errors.

Frequency	**States**		**Coefficient**	**SE**	**t-stat**	**p-value**
5 min	3 states	Intercept	0.02872	0.00538	5.33643	0.00000
		slope	4.40587	1.80102	2.44632	0.01738
	4 states	Intercept	0.03571	0.00494	7.23223	0.00000
		slope	3.93404	1.78874	2.19934	0.03172
	5 states	Intercept	0.04734	0.00399	11.86487	0.00000
		slope	1.68980	1.43897	1.17431	0.24491
10 min	3 states	Intercept	0.01056	0.00233	4.52421	0.00003
		slope	2.07871	0.91569	2.27009	0.02681
	4 states	Intercept	0.00145	0.00175	0.82488	0.41271
		slope	0.17092	0.43919	0.38916	0.69853
	5 states	Intercept	0.00409	0.00344	1.19164	0.23809
		slope	3.18249	1.40427	2.26630	0.02705
30 min	3 states	Intercept	0.00770	0.00157	4.89584	0.00001
		slope	0.33346	0.71038	0.46941	0.64048
	4 states	Intercept	0.00050	0.00137	0.36742	0.71460
		slope	-0.43771	0.78439	-0.55802	0.57890
	5 states	Intercept	0.00055	0.00134	0.41327	0.68088
		slope	0.76181	0.41203	1.84892	0.06940

Table 5.10: Jenson's Alpha (Discrete HMM, KOSPI50, 0.05% transaction cost)

The following tables compare the trading performance of the continuous Hidden Markov Model in KOSPI50 over the horizon of 1st November 2016 to 31st January 2017. The independent variable is the excess daily return of portfolio and the dependent variable is the excess daily market return of KOSPI. The adjusted daily 3-month Korea Interbank Offered Rate is used as a proxy to risk-free rate. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table E.7 and E.9 in appendix E. All standard errors are Heteroskedasticity-robust standard errors.

Frequency	**States**		**Coefficient**	**SE**	**t-stat**	**p-value**
5 min	3 states	Intercept	-0.00085	0.00062	-1.36643	0.17682
		slope	0.12419	0.11424	1.08710	0.28127
	4 states	Intercept	0.00045	0.00083	0.54786	0.58579
		Slope	0.17233	0.08384	2.05555	0.04411
	5 states	Intercept	0.00028	0.00155	0.18078	0.85714
		Slope	0.35100	0.19263	1.82211	0.07334
10 min	3 states	Intercept	-0.00099	0.00102	-0.97218	0.33480
		Slope	-0.20259	0.17676	-1.14613	0.25622
	4 states	Intercept	-0.00276	0.00090	-3.07415	0.00316
		Slope	-0.01946	0.09214	-0.21116	0.83347
	5 states	Intercept	-0.00166	0.00078	-2.14127	0.03626
		slope	-0.04272	0.05371	-0.79545	0.42944
30 min	3 states	Intercept	-0.00042	0.00036	-1.16637	0.24800
		slope	-0.17758	0.10339	-1.71762	0.09094
	4 states	Intercept	-0.00083	0.00084	-0.99094	0.32563
		slope	-0.16460	0.14618	-1.12596	0.26459
	5 states	Intercept	-0.00172	0.00095	-1.81365	0.07465
		slope	0.02500	0.12511	0.19981	0.84230

Table 5.11: Jenson's Alpha (Continuous HMM, SET50, 0.05% transaction cost)

The following tables compare the trading performance of the continuous Hidden Markov Model in SET50 over the horizon of 1st November 2016 to 31st January 2017. The prediction method of mean return (please see method I of section 4.4.4.3) is used for generating signal. The independent variable is the excess daily return of portfolio and the dependent variable is the excess daily market return of SET. The adjusted daily 3-month Bangkok Interbank Offered Rate is used as a proxy to risk-free rate. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table E.4 and E.6 in appendix E. All standard errors are Heteroskedasticity-robust standard errors.

Frequency	**States**		**Coefficient**	**SE**	**t-stat**	**p-value**
5 min	3 states	Intercept	-0.00076	0.00155	-0.48735	0.62779
		slope	2.35911	0.65034	3.62749	0.00059
	4 states	Intercept	-0.00081	0.00174	-0.46607	0.64286
		slope	2.18861	0.73122	2.99309	0.00401
	5 states	Intercept	0.00269	0.00213	1.26325	0.21139
		slope	2.56968	0.89342	2.87624	0.00556
10 min	3 states	Intercept	0.00052	0.00226	0.22982	0.81902
		slope	2.74045	0.94916	2.88723	0.00540
	4 states	Intercept	0.00076	0.00199	0.38166	0.70406
		slope	2.66650	0.83448	3.19541	0.00223
	5 states	Intercept	-0.00091	0.00221	-0.41270	0.68130
		slope	2.42797	0.92477	2.62547	0.01096
30 min	3 states	Intercept	0.00068	0.00191	0.35728	0.72214
		slope	2.58314	0.79908	3.23264	0.00199
	4 states	Intercept	-0.00027	0.00188	-0.14424	0.88580
		slope	2.32965	0.78866	2.95394	0.00448
	5 states	Intercept	0.00137	0.00183	0.74937	0.45656
		slope	2.76312	0.76797	3.59796	0.00065

Table 5.12: Jenson's Alpha (Continuous HMM, KOSPI50, 0.05% transaction cost)

The following tables compare the trading performance of the continuous Hidden Markov Model in KOSPI50 over the horizon of 1st November 2016 to 31st January 2017. The prediction method of mean return (please see method I of section 4.4.4.3) is used for generating signal. The independent variable is the excess daily return of portfolio and the dependent variable is the excess daily market return of KOSPI. The adjusted daily 3-month Korea Interbank Offered Rate is used as a proxy to risk-free rate. For this set of data, institutional investor's level of transaction cost (0.05%) is assumed for estimation of return, for other assumption of transaction cost, please see table E.10 and E.12 in appendix E. All standard errors are Heteroskedasticity-robust standard errors.

Frequency	**States**		**Coefficient**	**SE**	**t-stat**	**p-value**
5 min	3 states	Intercept	0.00299	0.00205	1.45514	0.15076
		slope	0.46458	0.22021	2.10973	0.03899
	4 states	Intercept	0.00038	0.00191	0.19750	0.84409
		slope	0.27108	0.28493	0.95138	0.34517
	5 states	Intercept	0.00048	0.00147	0.32415	0.74693
		slope	0.63291	0.27200	2.32686	0.02331
10 min	3 states	Intercept	-0.00022	0.00197	-0.11374	0.90982
		slope	0.34025	0.30512	1.11514	0.26916
	4 states	Intercept	0.00007	0.00197	0.03491	0.97227
		slope	0.19578	0.22615	0.86571	0.39004
	5 states	Intercept	0.00105	0.00230	0.45817	0.64846
		slope	0.37937	0.26423	1.43577	0.15618
30 min	3 states	Intercept	0.00116	0.00208	0.55657	0.57986
		slope	0.01775	0.21610	0.08216	0.93479
	4 states	Intercept	0.00065	0.00237	0.27600	0.78348
		slope	-0.05843	0.18161	-0.32174	0.74875
	5 states	Intercept	0.00168	0.00245	0.68566	0.49552
		slope	-0.11639	0.19171	-0.60712	0.54602

CHAPTER 6
CONCLUSION AND RECOMMENDATION

Inspired by the previous literature by Chordia et al (2005) that it is possible to predict stock price movement at intra-day level with order imbalance indicator, this study aims to capture the positive price movement resulted from the order imbalance state with both discrete and continuous Hidden Markov Models in a confident and consistent manner, which translates to high value of hit ratio and consistent number of signals generated. The later literature by Chrodia et al (2008) further reveals that the predictability of stock returns tends to disappear in a more liquid regime; with this evidence in mind, we set the scope of this study to Thai stock market and Korean stock market to assess the applicability of our models in markets with different liquidity.

The results from our models indicate different predictability and different profitability on different dimensions of frequency, market liquidity and types of models. In summary, the continuous models fail to generate a meaningful result, whereas the discrete models perform better in both profitability and predictability in a higher frequency where market liquidity is relatively lower. The following section discusses the implications of the evidences and recommendations for future study.

6.1 Performance of the discrete and continuous models

In this study, we propose one approach for the discrete models (predict by probability) and two approaches (predict by mean return or probability) for the continuous models to generate the signals to predict the market movement.

For all frequencies, all models with different number of states, the results from the continuous models show lack of predictability; as a result, the trading strategy is not able to generate any abnormal return. On the other hand, the discrete models are able to achieve varied degrees of hit ratio on different frequency and market liquidity and as a result the profitability of the strategy is highly dependent on predictability of the models.

We come to the conclusion that the reason of continuous models fail to generate meaningful result in both approach is due to the assumption of distribution. As discussed in section 4, due to the non-normality of intra-day data, this study incorporates the 3 components Gaussian Mixture Model as the distribution function to describe the observable emission of the Hidden Markov Models. However, as the result in appendix C has shown, the BIC score of the GMM models are better, but not remarkably better. In light of this concern, the estimated models are unable to obtain a suitable distribution that best describes the emission. In contrast, the discrete models do not require explicit assumption but requires an appropriate method for discretizing the data. In this study, we have presented a trivial method for discretizing the stock price movement and order imbalance ratio, but there is still lot of rooms for improvement which might benefit the model.

Due to the fact that the continuous models fail to generate meaningful result, in the following section, we focus the discussion of the study on the result of discrete models.

6.2 Implication on the performance of the models in different frequencies

As discussed in section 5, the predictability of the models decreases as we lower the frequency (from 5 minutes to 30 minutes). The decrease in predictability takes in two forms in this study; first as the frequency decreases, the observed average hit ratio decreases. Second, as the frequency decreases, the models are less consistent and confident in generating signals. For instance, moving from 5 minute to 10 minute, the total number of signals generated from the models decrease at a factor of 5 instead of the expected number of 2.

The decrease in confidence of the models in making a prediction seems to be consistent with the previous literature by Chordia et al (2005). The predictability of stock price tends to disappear over a short horizon due the price are already adjusted to the new information. Therefore, as time passes, the stock price movement tends to random walk. The models are unable to make prediction because no clear patterns are observed from the data.

Expectedly, due to better predictability in high frequency data, the trading strategy gain the highest abnormal return and Sharpe’s Ratio for the highest frequency case in the Thai market. Due to the lower predictability of the models in the Korean market, the strategy is only able to generate abnormal return when there is no transaction presents. We further discuss the possible explanation on why the models achieve lower predictability in the Korean market in the following section.

6.3 Implication on the performance of the models in different markets

Detailed from the study by Chordia et al (2008), the market liquidity enhances the speed of convergence to market efficiency. As a result, in a more liquid market, the predictability of stocks tends to disappear. Therefore, to assess the applicability of the models under different liquidity environment, we test our strategy in both Korean and Thai stock market, where the market liquidity in the Korean market is relatively higher.

The result seems to be consistent with the previous literature, the predictability of the models are relatively better in terms of both hit ratio and number of signals generated in Thai stock market in comparison to Korean stock market given that the daily trading time in Korean stock market is longer than Thailand; and due to the lower predictability, even at the highest frequency, the strategy is not able to generate abnormal return at institutional level of transaction cost and achieves abnormal return only at case when transaction cost does not exist.

One logical explanation on why market efficiency is enhanced in a more liquid regime is the narrower bid-ask spread. With low bid-ask spread, the barrier for investors to take advantage of momentarily mispricing of financial assets; and as a result, the speed adjustment of information into price is enhanced.

6.4 Recommendation for further study

Based on the observations from this study, the discrete Hidden Markov Model is the more appropriate model than the continuous version due to the benefit of no requirement on assumption of distribution. However, there are rooms for improving the methods of discretizing data; in this study, we propose a simple method for discretizing with order imbalance ratio and stock price movement. Due the process of forming the order imbalance ratio, the information of absolute size of order imbalance is lost; the relevance of the absolute size of order imbalance remains untouched.

Therefore, it is recommended for future study to formulate a method to incorporate the absolute size of order imbalance into the discretization process.

In addition to the improvement on the models, future study can also consider increasing the frequency of data. For this study, we set the highest frequency of data to be 5 minutes in order to avoid occurrences of no trade. However, for purpose of making gain from trading, the pursuer of opportunity should aim for higher frequency data.

REFERENCES

1. Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order Imbalance, liquidity and market returns. Journal of Financial Economics, 3-28.

2. Chordia, T., Roll, R., & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. Journal of Financial Economics, 271-292.

3. Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. Journal of Financial Economics, 249-268.

4. Chuin Ching, L., & Tak Kuen, S. (2010). A hidden Markov regime-switching model for option valuation. Insurance: Mathematics and Economics, 374-384.

5. Fama, F. (1970). Efficient Capital Market: A Review of Theory and Empirical Work. Journal of Finance, 383-417.

6. Giacomo, G., Mark, D., & Crowder, M. (2005). Analysis of Default Data Using Hidden Markov Models. Quantitative Finance, 27-34.

7. Haipeng, X., Ning, S., & Ying, C. (2012). Credit rating dynamics in the presence of unknown structural breaks. Journal of Banking & Finance, 78-89.

8. Hassan, R. (2005). Stock market forecasting using hidden Markov model: a new approach. 5th International Conference on Intelligent Systems Design and Applications, 192-196.

9. Hassan, R. (2009). A combination of hidden Markov model and fuzzy model for stock marke forecasting. Neurocomputing, 3439–3446.

10. Hassan, R. (2013). A HMM-basedadaptivefuzzyinferencesystemforstock marketforecasting. Neurocomputing, 10-25.

11. Korolkiewicz, M., & Elliott, R. (2008). A hiddenMarkovmodelofcreditquality. Journal of Economic Dynamics & Control, 3807-3819.

12. Patrik, I., & Conny, J. (2008). Algorithmic Trading: Hidden Markov Models on Foreign Exchange Data. Master Thesis, Department of Mathematics, Linkopings University.

13. Robert, E., & Tak Kuen, S. (2013). Option Pricing and Filtering with Hidden Markov-Modulated Pure-Jump Processes. Applied Mathematical Finance, 1-25.

14. Satish, R., & Jerry, H. (2010). Analysis of Hidden Markov Models and Support Vector Machines in Financial Applications. Berkeley: University of California .

15. Shangzhen, L., & Xudong, Z. (2014). An optimal investment model with Markov-Driven Volatilities. Quantitative Finance, 1651-1661.

16. Shen, D. (2015, May 27). Order Imbalance Based Strategy in High Frequency Trading. Master Thesis of Linacre College.

17. Sittipong, S. (2012). Forecasting SET50 Index With Neuro-Fuzzy System And Technical Analysis. Independent Study, Thammasat University.

18. Tenyakov, A. (2014). Estimation of Hidden Markov Models and Their Applications in Finance. Electronic Thesis and Dissertation Repository, 2348.

19. Thapanun, P. (2013). Forecasting Stock Indices Movement Using Hybrid Model: A Comparison of Traditional And Machine Learning Approaches. Independent Study, Thammasat University.

APPENDICES

APPENDIX A
SUMMARY STATISTICS

Table A.1: Summary statistics of SET50: return of 5 Minute data

Ticker	μ	σ	Skewness	kurtosis
ADVANC.BK	0.00000	0.00287	-0.57476	35.13078
BANPU.BK	0.00004	0.00448	-0.19062	13.17991
BCP.BK	0.00003	0.00504	-0.00774	4.37384
CPF.BK	-0.00003	0.00607	-0.19439	6.49320
DTAC.BK	0.00006	0.00577	-0.43297	14.74661
IRPC.BK	0.00001	0.00420	-0.58875	15.65829
IVL.BK	0.00005	0.00575	-0.40375	10.40947
PTTEP.BK	0.00004	0.00293	0.20801	22.99403
TCAP.BK	0.00003	0.00422	0.15091	4.66762
TRUE.BK	-0.00002	0.00538	-0.01786	13.59031

Table A.2: Summary statistics of SET50: Order Imbalance ratio of 5 Minute data

Ticker	μ	σ	Skewness	kurtosis
ADVANC.BK	0.53423	0.15950	-0.17502	2.78032
BANPU.BK	0.51051	0.16400	-0.08301	2.52451
BCP.BK	0.50299	0.20999	0.09503	2.33031
CPF.BK	0.53974	0.16877	-0.34334	2.84258
DTAC.BK	0.52986	0.18704	-0.10703	2.46939
IRPC.BK	0.53441	0.19070	-0.22314	2.49436
IVL.BK	0.52873	0.17074	-0.16115	2.63426
PTTEP.BK	0.51626	0.15895	-0.12123	2.67707
TCAP.BK	0.50426	0.20520	-0.06756	2.44605
TRUE.BK	0.53524	0.18212	-0.12667	2.52965

Table A.3: Summary statistics of SET50: Correlation between return and OIR of 5 minute data

Ticker	Pearson	P-Value $H_0: \rho = 0$
ADVANC.BK	-0.02038	0.17213
BANPU.BK	-0.07691	0.00000
BCP.BK	-0.05191	0.00050
CPF.BK	-0.01930	0.19581
DTAC.BK	-0.06847	0.00000
IRPC.BK	-0.03166	0.03386
IVL.BK	-0.07370	0.00000
PTTEP.BK	-0.03465	0.02023
TCAP.BK	-0.04752	0.00144
TRUE.BK	-0.03537	0.01775

Table A.4: Summary statistics of SET50: return of 10 Minute data

Ticker	μ	σ	Skewness	kurtosis
ADVANC.BK	0.00000	0.00334	-0.66030	34.82407
BANPU.BK	0.00008	0.00548	-0.13297	11.77086
BCP.BK	0.00005	0.00560	0.01227	4.49157
CPF.BK	-0.00005	0.00668	-0.26767	7.91537
DTAC.BK	0.00012	0.00695	-0.35091	13.43625
IRPC.BK	0.00002	0.00472	-0.72947	16.82524
IVL.BK	0.00008	0.00664	-0.32751	11.68384
PTTEP.BK	0.00008	0.00362	0.32464	19.40692
TCAP.BK	0.00006	0.00474	0.13991	4.87832
TRUE.BK	-0.00003	0.00619	0.20240	14.99984

Table A.5: Summary statistics of SET50: Order Imbalance ratio of 10 Minute data

Ticker	μ	σ	Skewness	kurtosis
ADVANC.BK	0.53438	0.14970	-0.18953	2.85399
BANPU.BK	0.50927	0.15433	-0.10084	2.64586
BCP.BK	0.50377	0.19854	0.11529	2.34607
CPF.BK	0.53938	0.16356	-0.32750	2.80337
DTAC.BK	0.52755	0.17513	-0.10444	2.54849
IRPC.BK	0.53286	0.18249	-0.22121	2.53322
IVL.BK	0.52529	0.16552	-0.16730	2.68864
PTTEP.BK	0.51680	0.14493	-0.14370	2.74825
TCAP.BK	0.50478	0.19425	-0.07829	2.50576
TRUE.BK	0.53646	0.17483	-0.13259	2.59712

Table A.6: Summary statistics of SET50: Correlation between return and OIR of 10 minute data

Ticker	Pearson	P-Value $H_0: \rho = 0$
ADVANC.BK	0.00345	0.86444
BANPU.BK	-0.06756	0.00081
BCP.BK	-0.03944	0.05075
CPF.BK	-0.00863	0.66914
DTAC.BK	-0.06391	0.00154
IRPC.BK	-0.03483	0.08456
IVL.BK	-0.07383	0.00025
PTTEP.BK	0.01447	0.47356
TCAP.BK	-0.04952	0.01415
TRUE.BK	-0.02523	0.21155

Table A.7: Summary statistics of SET50: return of 30 Minute data

Ticker	μ	σ	Skewness	kurtosis
ADVANC.BK	0.00000	0.00443	-0.44794	20.43120
BANPU.BK	0.00024	0.00755	-0.15494	7.12534
BCP.BK	0.00007	0.00700	-0.01969	4.51733
CPF.BK	0.00000	0.00796	-0.22049	7.19404
DTAC.BK	0.00006	0.00892	-0.03094	10.07861
IRPC.BK	0.00003	0.00532	-0.65818	15.35638
IVL.BK	0.00005	0.00894	-0.02576	10.01281
PTTEP.BK	0.00011	0.00502	0.28247	12.95480
TCAP.BK	0.00018	0.00600	-0.02493	4.51603
TRUE.BK	-0.00005	0.00786	0.55802	11.93959

Table A.8: Summary statistics of SET50: Order Imbalance ratio of 30 Minute data

Ticker	μ	σ	Skewness	kurtosis
ADVANC.BK	0.52782	0.13041	-0.28024	3.11894
BANPU.BK	0.49930	0.13647	-0.26399	2.89891
BCP.BK	0.51724	0.17104	-0.02045	2.45173
CPF.BK	0.52705	0.14903	-0.32937	2.82542
DTAC.BK	0.52311	0.15553	-0.15458	2.76485
IRPC.BK	0.52004	0.15725	-0.19218	2.70322
IVL.BK	0.50451	0.15014	-0.32200	2.97702
PTTEP.BK	0.51068	0.12828	-0.30303	3.00454
TCAP.BK	0.51779	0.18016	-0.06109	2.53002
TRUE.BK	0.52344	0.16547	-0.08876	2.74810

Table A.9: Summary statistics of SET50: Correlation between return and OIR of 30 minute data

Ticker	Pearson	P-Value $H_0: \rho = 0$
ADVANC.BK	0.03924	0.08870
BANPU.BK	-0.02826	0.22022
BCP.BK	-0.01635	0.47832
CPF.BK	0.00927	0.68765
DTAC.BK	-0.02592	0.26096
IRPC.BK	-0.00536	0.81635
IVL.BK	-0.03805	0.09885
PTTEP.BK	0.04879	0.03424
TCAP.BK	-0.02856	0.21546
TRUE.BK	-0.04001	0.08261

Table A.10: Summary statistics of KOSPI50: return of 5 Minute data

Ticker	μ	σ	Skewness	kurtosis
034220.KS	0.00001	0.00262	-0.19991	24.20587
066570.KS	0.00003	0.00212	0.61680	15.73020
051910.KS	0.00002	0.00237	-1.27951	42.47450
005490.KS	0.00003	0.00254	0.74383	23.81689
006400.KS	0.00003	0.00282	-0.35090	25.74632
009150.KS	0.00002	0.00209	2.92288	76.95801
010140.KS	0.00002	0.00289	1.00025	17.88761
000880.KS	0.00000	0.00215	0.88572	17.47675
000720.KS	0.00002	0.00243	-0.29044	20.36857
009540.KS	0.00000	0.00354	3.27655	96.11003

Table A.11: Summary statistics of KOSPI50: Order Imbalance ratio of 5 Minute data

Ticker	μ	σ	Skewness	kurtosis
034220.KS	0.50278	0.18441	0.04378	2.44974
066570.KS	0.45617	0.16836	0.09041	2.55937
051910.KS	0.48191	0.18564	0.07451	2.49309
005490.KS	0.48804	0.17662	0.11903	2.58783
006400.KS	0.46574	0.19397	0.09358	2.34988
009150.KS	0.47322	0.19225	0.05850	2.33684
010140.KS	0.46239	0.18132	0.16428	2.46574
000880.KS	0.49894	0.19414	-0.01829	2.31689
000720.KS	0.46686	0.18972	0.02007	2.38270
009540.KS	0.48016	0.20594	0.06555	2.26023

Table A.12: Summary statistics of KOSPI50: Correlation between return and OIR of 5 minute data

Ticker	Pearson	P-Value $H_0: \rho = 0$
034220.KS	0.00599	0.63066
066570.KS	0.04209	0.00072
051910.KS	-0.02268	0.06862
005490.KS	0.02734	0.02818
006400.KS	0.01722	0.16683
009150.KS	-0.01175	0.34565
010140.KS	-0.02236	0.07259
000880.KS	0.01720	0.16720
000720.KS	0.00852	0.49387
009540.KS	-0.04025	0.00123

Table A.13: Summary statistics of KOSPI50: return of 10 Minute data

Ticker	μ	σ	Skewness	kurtosis
034220.KS	0.00002	0.00347	0.21555	16.09205
066570.KS	0.00006	0.00280	0.57462	17.41173
051910.KS	0.00004	0.00304	-0.94218	26.72175
005490.KS	0.00006	0.00339	0.49067	17.94879
006400.KS	0.00006	0.00346	0.55281	23.13920
009150.KS	0.00004	0.00281	2.01350	38.86501
010140.KS	0.00004	0.00375	1.03344	18.98968
000880.KS	0.00000	0.00288	0.72423	16.12458
000720.KS	0.00004	0.00337	0.37581	17.74160
009540.KS	0.00001	0.00436	3.67801	84.59788

Table A.14: Summary statistics of KOSPI50: Order Imbalance ratio of 10 Minute data

Ticker	μ	σ	Skewness	kurtosis
034220.KS	0.50266	0.16906	0.04839	2.62821
066570.KS	0.45578	0.15294	0.07045	2.62186
051910.KS	0.48259	0.17052	0.07871	2.59031
005490.KS	0.48780	0.15991	0.11255	2.68708
006400.KS	0.46612	0.18076	0.09743	2.40920
009150.KS	0.47433	0.17877	0.04428	2.40740
010140.KS	0.46272	0.16554	0.13564	2.56070
000880.KS	0.49997	0.18061	-0.02333	2.35060
000720.KS	0.46531	0.17699	-0.01198	2.48084
009540.KS	0.48107	0.18810	0.06867	2.38551

Table A.15: Summary statistics of KOSPI50: Correlation between return and OIR of 10 minute data

Ticker	Pearson	P-Value $H_0: \rho = 0$
034220.KS	-0.04382	0.01285
066570.KS	0.02451	0.16413
051910.KS	-0.05826	0.00094
005490.KS	0.00863	0.62437
006400.KS	0.01239	0.48204
009150.KS	-0.03979	0.02387
010140.KS	-0.04344	0.01364
000880.KS	-0.01079	0.54018
000720.KS	-0.02499	0.15601
009540.KS	-0.06882	0.00009

Table A.16: Summary statistics of KOSPI50: return of 30 Minute data

Ticker	μ	σ	Skewness	kurtosis
034220.KS	0.00005	0.00591	0.03757	16.46221
066570.KS	-0.00002	0.00480	-1.59037	46.40468
051910.KS	-0.00007	0.00557	-3.16086	71.95813
005490.KS	0.00009	0.00579	0.41158	15.48772
006400.KS	0.00007	0.00643	0.60117	19.55338
009150.KS	0.00000	0.00492	1.09920	35.93384
010140.KS	0.00004	0.00722	1.18305	55.58516
000880.KS	0.00001	0.00512	0.14878	12.67493
000720.KS	0.00005	0.00566	0.12278	13.29544
009540.KS	0.00012	0.00700	0.69218	15.43215

Table A.17: Summary statistics of KOSPI50: Order Imbalance ratio of 30 Minute data

Ticker	μ	σ	Skewness	kurtosis
034220.KS	0.49235	0.14952	0.18718	3.05423
066570.KS	0.45883	0.13771	0.12219	3.07626
051910.KS	0.50275	0.14436	0.00031	2.83987
005490.KS	0.49712	0.14215	0.06058	3.10312
006400.KS	0.46376	0.16116	0.10762	2.78443
009150.KS	0.49050	0.15978	-0.06891	2.64483
010140.KS	0.47290	0.14246	0.02502	2.81911
000880.KS	0.51241	0.15402	-0.12731	2.80382
000720.KS	0.49857	0.15619	-0.18739	2.99526
009540.KS	0.47630	0.16320	0.05299	2.70564

Table A.18: Summary statistics of KOSPI50: Correlation between return and OIR of 30 minute data

Ticker	Pearson	P-Value $H_0: \rho = 0$
034220.KS	-0.08661	0.00001
066570.KS	0.01994	0.31665
051910.KS	-0.16167	0.00000
005490.KS	-0.11902	0.00000
006400.KS	-0.05228	0.00861
009150.KS	-0.13250	0.00000
010140.KS	-0.08867	0.00001
000880.KS	-0.10166	0.00000
000720.KS	-0.12355	0.00000
009540.KS	-0.14270	0.00000

APPENDIX B
NORMALITY TEST

Table B.1: Normality Test on SET50 5 Minute Data

	Shapiro-Wilk Normality Test H_0: Distribution is normal P-Value		Henze-Zirkler MVN Test Lagged return, % change in volume H_0: Distribution is normal	
Ticker	**return**	**OIR**	**HZ statistics**	**P-Value**
ADVANC.BK	0.00000	0.00000	113.70910	0.00000
BANPU.BK	0.00000	0.00000	124.11596	0.00000
BCP.BK	0.00000	0.00000	203.93894	0.00000
CPF.BK	0.00000	0.00000	168.16300	0.00000
DTAC.BK	0.00000	0.00000	125.82064	0.00000
IRPC.BK	0.00000	0.00000	134.05994	0.00000
IVL.BK	0.00000	0.00000	158.12520	0.00000
PTTEP.BK	0.00000	0.00000	72.37816	0.00000
TCAP.BK	0.00000	0.00000	164.29306	0.00000
TRUE.BK	0.00000	0.00000	159.59546	0.00000

Table B.2: Normality Test on SET50 10 Minute Data

	Shapiro-Wilk Normality Test H_0: Distribution is normal P-Value		Henze-Zirkler MVN Test Lagged return, % change in volume H_0: Distribution is normal	
Ticker	**return**	**OIR**	**HZ statistics**	**P-Value**
ADVANC.BK	0.00000	0.00002	32.21822	0.00000
BANPU.BK	0.00000	0.00018	30.39697	0.00000
BCP.BK	0.00000	0.00000	83.23221	0.00000
CPF.BK	0.00000	0.00000	69.37168	0.00000
DTAC.BK	0.00000	0.00000	34.55117	0.00000
IRPC.BK	0.00000	0.00000	47.74824	0.00000
IVL.BK	0.00000	0.00000	55.97671	0.00000
PTTEP.BK	0.00000	0.00002	14.91344	0.00000
TCAP.BK	0.00000	0.00000	61.62666	0.00000
TRUE.BK	0.00000	0.00000	58.87801	0.00000

Table B.3: Normality Test on SET50 30 Minute Data

	Shapiro-Wilk Normality Test H_0: Distribution is normal P-Value		Henze-Zirkler MVN Test Lagged return, % change in volume H_0: Distribution is normal	
Ticker	**return**	**OIR**	**HZ statistics**	**P-Value**
ADVANC.BK	0.00000	0.00001	10.23396	0.00000
BANPU.BK	0.00000	0.00000	9.53336	0.00000
BCP.BK	0.00000	0.00000	29.37754	0.00000
CPF.BK	0.00000	0.00000	27.46295	0.00000
DTAC.BK	0.00000	0.00012	12.21123	0.00000
IRPC.BK	0.00000	0.00002	16.79580	0.00000
IVL.BK	0.00000	0.00000	14.34325	0.00000
PTTEP.BK	0.00000	0.00000	10.37672	0.00000
TCAP.BK	0.00000	0.00001	19.70027	0.00000
TRUE.BK	0.00000	0.00241	13.63457	0.00000

Table B.4: Normality Test on KOSPI50 5 Minute Data

	Shapiro-Wilk Normality Test H_0: Distribution is normal P-Value		Henze-Zirkler MVN Test Lagged return, % change in volume H_0: Distribution is normal	
Ticker	**return**	**OIR**	**HZ statistics**	**P-Value**
034220.KS	0.00000	0.00000	34.13249	0.00000
066570.KS	0.00000	0.00000	45.54455	0.00000
051910.KS	0.00000	0.00000	74.92932	0.00000
005490.KS	0.00000	0.00000	56.48314	0.00000
006400.KS	0.00000	0.00000	135.38457	0.00000
009150.KS	0.00000	0.00000	56.21242	0.00000
010140.KS	0.00000	0.00000	82.57328	0.00000
000880.KS	0.00000	0.00000	27.52713	0.00000
000720.KS	0.00000	0.00000	36.53064	0.00000
009540.KS	0.00000	0.00000	102.42044	0.00000

Table B.5: Normality Test on KOSPI50 10 Minute Data

	Shapiro-Wilk Normality Test H_0: Distribution is normal P-Value		Henze-Zirkler MVN Test Lagged return, % change in volume H_0: Distribution is normal	
Ticker	**return**	**OIR**	**HZ statistics**	**P-Value**
034220.KS	0.00000	0.00001	16.73531	0.00000
066570.KS	0.00000	0.00000	21.54032	0.00000
051910.KS	0.00000	0.00000	18.54945	0.00000
005490.KS	0.00000	0.00001	19.82128	0.00000
006400.KS	0.00000	0.00000	59.74719	0.00000
009150.KS	0.00000	0.00000	33.33448	0.00000
010140.KS	0.00000	0.00000	34.02990	0.00000
000880.KS	0.00000	0.00000	14.78238	0.00000
000720.KS	0.00000	0.00000	24.10563	0.00000
009540.KS	0.00000	0.00000	21.07727	0.00000

Table B.6: Normality Test on KOSPI50 30 Minute Data

	Shapiro-Wilk Normality Test H_0: Distribution is normal P-Value		Henze-Zirkler MVN Test Lagged return, % change in volume H_0: Distribution is normal	
Ticker	**Return**	**OIR**	**HZ statistics**	**P-Value**
034220.KS	0.00000	0.00031	20.86701	0.00000
066570.KS	0.00000	0.00952	22.75294	0.00000
051910.KS	0.00000	0.30366	30.41895	0.00000
005490.KS	0.00000	0.03579	20.67882	0.00000
006400.KS	0.00000	0.00051	25.69704	0.00000
009150.KS	0.00000	0.00016	27.13374	0.00000
010140.KS	0.00000	0.07027	30.64125	0.00000
000880.KS	0.00000	0.00408	15.82108	0.00000
000720.KS	0.00000	0.00000	23.49415	0.00000
009540.KS	0.00000	0.00670	12.16406	0.00000

APPENDIX C
GMM FITTING BY USING EXPECTATION MAXIMIZATION ALGORITHM

Table C.1: GMM Fitting: SET50 5 minute data

Ticker	Number of Components	Negative Loglikelihood	AIC	BIC
ADVANC.BK	1	-21239.9	-42469.9	-42437.8
	2	-21465	-42908	-42837.5
	3	-21482.2	-42930.3	-42821.4
	4	-21494.4	-42942.8	-42795.3
	5	-21483.2	-42908.3	-42722.4
BANPU.BK	1	-19522.4	-39034.8	-39002.7
	2	-19685.9	-39349.9	-39279.3
	3	-19686.2	-39338.4	-39229.4
	4	-19728.2	-39410.4	-39263
	5	-19732.8	-39407.5	-39221.6
BCP.BK	1	-17920.4	-35830.8	-35798.7
	2	-18012.6	-36003.1	-35932.6
	3	-18045.2	-36056.4	-35947.4
	4	-18046.8	-36047.5	-35900.1
	5	-18079.1	-36100.1	-35914.2
CPF.BK	1	-18117.2	-36224.5	-36192.4
	2	-18180.1	-36338.1	-36267.6
	3	-18257.1	-36480.2	-36371.3
	4	-18260	-36474	-36326.6
	5	-18262.6	-36467.1	-36281.2
DTAC.BK	1	-17876.7	-35743.4	-35711.3
	2	-18092.6	-36163.2	-36092.6
	3	-18138.3	-36242.6	-36133.6
	4	-18155.4	-36264.7	-36117.3
	5	-18150.2	-36242.5	-36056.6
IRPC.BK	1	-19090.7	-38171.4	-38139.3
	2	-19244.9	-38467.7	-38397.2
	3	-19307.1	-38580.1	-38471.1
	4	-19341.1	-38636.1	-38488.7
	5	-19321.3	-38584.6	-38398.7
IVL.BK	1	-18311.2	-36612.5	-36580.4
	2	-18356.5	-36691	-36620.5
	3	-18471.5	-36909	-36800
	4	-18471.6	-36897.1	-36749.7
	5	-18476.5	-36895	-36709.1
PTTEP.BK	1	-21199.6	-42389.1	-42357.1
	2	-21426.1	-42830.1	-42759.6
	3	-21441.2	-42848.3	-42739.3
	4	-21450.5	-42854.9	-42707.5
	5	-21452.9	-42847.7	-42661.8
TCAP.BK	1	-18748.8	-37487.7	-37455.6
	2	-18787.7	-37553.4	-37482.9
	3	-18785.6	-37537.2	-37428.2
	4	-18785.8	-37525.7	-37378.3
	5	-18787.6	-37517.3	-37331.4
TRUE.BK	1	-18290.4	-36570.8	-36538.8
	2	-18373.1	-36724.1	-36653.6
	3	-18523.1	-37012.2	-36903.2
	4	-18548.3	-37050.6	-36903.1
	5	-18525.3	-36992.6	-36806.7

Table C.2: GMM Fitting: SET50 10 minute data

Ticker	Number of Components	Negative Loglikelihood	AIC	BIC
ADVANC.BK	1	-11484.7	-22959.3	-22930.3
	2	-11673.4	-23324.9	-23261
	3	-11681.6	-23329.2	-23230.5
	4	-11683.5	-23321	-23187.4
	5	-11691.5	-23324.9	-23156.5
BANPU.BK	1	-10360.1	-20710.2	-20681.2
	2	-10475.4	-20928.9	-20865
	3	-10489.7	-20945.4	-20846.7
	4	-10482.4	-20918.9	-20785.4
	5	-10503.3	-20948.5	-20780.2
BCP.BK	1	-9685.81	-19361.6	-19332.6
	2	-9720.29	-19418.6	-19354.7
	3	-9735.09	-19436.2	-19337.5
	4	-9770.25	-19494.5	-19361
	5	-9773.82	-19489.6	-19321.3
CPF.BK	1	-9748.26	-19486.5	-19457.5
	2	-9827.03	-19632.1	-19568.2
	3	-9858.74	-19683.5	-19584.8
	4	-9861.8	-19677.6	-19544.1
	5	-9865.92	-19673.8	-19505.5
DTAC.BK	1	-9491.92	-18973.8	-18944.8
	2	-9657.4	-19292.8	-19228.9
	3	-9675.97	-19317.9	-19219.3
	4	-9691.68	-19337.4	-19203.8
	5	-9695.84	-19333.7	-19165.3
IRPC.BK	1	-10281.7	-20553.4	-20524.4
	2	-10407.2	-20792.3	-20728.5
	3	-10438.1	-20842.3	-20743.6
	4	-10447.7	-20849.5	-20716
	5	-10468.1	-20878.1	-20709.8
IVL.BK	1	-9740.21	-19470.4	-19441.4
	2	-9864.13	-19706.3	-19642.4
	3	-9880.37	-19726.7	-19628.1
	4	-9885.03	-19724.1	-19590.5
	5	-9904.92	-19751.8	-19583.5
PTTEP.BK	1	-11408.9	-22807.8	-22778.7
	2	-11595.1	-23168.2	-23104.3
	3	-11596.8	-23159.6	-23060.9
	4	-11607.7	-23169.4	-23035.9
	5	-11613.6	-23169.2	-23000.9
TCAP.BK	1	-10120.4	-20230.8	-20201.8
	2	-10138.5	-20255	-20191.2
	3	-10157	-20280.1	-20181.4
	4	-10156.5	-20267	-20133.5
	5	-10164.6	-20271.1	-20102.8
TRUE.BK	1	-9764.92	-19519.8	-19490.8
	2	-9964.37	-19906.7	-19842.9
	3	-9995.69	-19957.4	-19858.7
	4	-10008.6	-19971.1	-19837.6
	5	-10001.4	-19944.7	-19776.4

Table C.3: GMM Fitting: SET50 30 minute data

Ticker	Number of Components	Negative Loglikelihood	AIC	BIC
ADVANC.BK	1	-8633.99	-17258	-17230.3
	2	-8808.63	-17595.3	-17534.3
	3	-8812.45	-17590.9	-17496.7
	4	-8816.09	-17586.2	-17458.7
	5	-8826.82	-17595.6	-17435
BANPU.BK	1	-7600.9	-15191.8	-15164.1
	2	-7675.8	-15329.6	-15268.7
	3	-7695.97	-15357.9	-15263.7
	4	-7715.43	-15384.9	-15257.4
	5	-7720.67	-15383.3	-15222.6
BCP.BK	1	-7313.35	-14616.7	-14589
	2	-7344.74	-14667.5	-14606.5
	3	-7347.34	-14660.7	-14566.5
	4	-7375.27	-14704.5	-14577.1
	5	-7379.15	-14700.3	-14539.6
CPF.BK	1	-7337.99	-14666	-14638.3
	2	-7399.81	-14777.6	-14716.7
	3	-7434.77	-14835.5	-14741.3
	4	-7433.88	-14821.8	-14694.3
	5	-7445.01	-14832	-14671.3
DTAC.BK	1	-7043.28	-14076.6	-14048.9
	2	-7203.84	-14385.7	-14324.7
	3	-7221.75	-14409.5	-14315.3
	4	-7228.5	-14411	-14283.6
	5	-7243.61	-14429.2	-14268.5
IRPC.BK	1	-7966.22	-15922.4	-15894.7
	2	-8097.68	-16173.4	-16112.4
	3	-8110.47	-16186.9	-16092.7
	4	-8110.88	-16175.8	-16048.3
	5	-8137.3	-16216.6	-16055.9
IVL.BK	1	-7107.95	-14205.9	-14178.2
	2	-7231.37	-14440.7	-14379.8
	3	-7266.78	-14499.6	-14405.4
	4	-7257.59	-14469.2	-14341.7
	5	-7290.95	-14523.9	-14363.2
PTTEP.BK	1	-8453.91	-16897.8	-16870.1
	2	-8633.65	-17245.3	-17184.3
	3	-8640.43	-17246.9	-17152.7
	4	-8649.51	-17253	-17125.6
	5	-8665.51	-17273	-17112.3
TCAP.BK	1	-7496.7	-14983.4	-14955.7
	2	-7507.07	-14992.1	-14931.2
	3	-7528.58	-15023.2	-14929
	4	-7528.36	-15010.7	-14883.3
	5	-7541.85	-15025.7	-14865
TRUE.BK	1	-7162.46	-14314.9	-14287.2
	2	-7361.54	-14701.1	-14640.1
	3	-7361.69	-14689.4	-14595.2
	4	-7399.95	-14753.9	-14626.5
	5	-7409.91	-14761.8	-14601.1

Table C.4: GMM Fitting: KOSPI50 5 minute data

Ticker	Number of Components	Negative Loglikelihood	AIC	BIC
034220.KS	1	-29937.7	-59865.4	-59831.6
	2	-30162.5	-60302.9	-60228.4
	3	-30214.2	-60394.4	-60279.3
	4	-30215.9	-60385.9	-60230.1
	5	-30245.7	-60433.3	-60237
066570.KS	1	-31341.4	-62672.8	-62639
	2	-31374.8	-62727.7	-62653.2
	3	-31382.1	-62730.3	-62615.2
	4	-31438.5	-62831.1	-62675.3
	5	-31472.6	-62887.2	-62690.8
051910.KS	1	-30305.8	-60601.6	-60567.7
	2	-30533.6	-61045.2	-60970.7
	3	-30577.5	-61121.1	-61006
	4	-30575.8	-61105.7	-60949.9
	5	-30579.6	-61101.2	-60904.8
005490.KS	1	-30355.2	-60700.5	-60666.6
	2	-30646.2	-61270.3	-61195.8
	3	-30683.6	-61333.2	-61218
	4	-30684.1	-61322.1	-61166.4
	5	-30685	-61312	-61115.6
006400.KS	1	-29306.7	-58603.5	-58569.6
	2	-29534.8	-59047.6	-58973.1
	3	-29620.4	-59206.9	-59091.8
	4	-29660.6	-59275.3	-59119.5
	5	-29671	-59284	-59087.6
009150.KS	1	-30530.1	-61050.2	-61016.3
	2	-30747.5	-61472.9	-61398.4
	3	-30841	-61648.1	-61532.9
	4	-30854.4	-61662.9	-61507.1
	5	-30835.2	-61612.4	-61416.1
010140.KS	1	-29631.9	-59253.8	-59220
	2	-29721.6	-59421.1	-59346.6
	3	-29980.3	-59926.6	-59811.5
	4	-29992.5	-59939.1	-59783.3
	5	-29966.1	-59874.3	-59677.9
000880.KS	1	-30360.4	-60710.8	-60677
	2	-30360.4	-60698.8	-60624.3
	3	-30547.8	-61061.6	-60946.5
	4	-30570.8	-61095.5	-60939.8
	5	-30548.1	-61038.3	-60841.9
000720.KS	1	-30064.1	-60118.2	-60084.4
	2	-30136.1	-60250.2	-60175.7
	3	-30278.8	-60523.6	-60408.5
	4	-30296	-60546	-60390.3
	5	-30295.3	-60532.7	-60336.3
009540.KS	1	-27824.1	-55638.3	-55604.4
	2	-28228.6	-56435.1	-56360.6
	3	-28335.6	-56637.1	-56522
	4	-28342.9	-56639.9	-56484.1
	5	-28362.8	-56667.6	-56471.2

Table C.5: GMM Fitting: KOSPI50 10 minute data

Ticker	Number of Components	Negative Loglikelihood	AIC	BIC
034220.KS	1	-14598.7	-29187.3	-29156.9
	2	-14796.2	-29570.4	-29503.5
	3	-14807.6	-29581.3	-29478
	4	-14811.6	-29577.1	-29437.3
	5	-14820.5	-29583	-29406.8
066570.KS	1	-15429.8	-30849.7	-30819.3
	2	-15575.4	-31128.9	-31062
	3	-15588.5	-31143	-31039.7
	4	-15589.1	-31132.2	-30992.4
	5	-15588.7	-31119.3	-30943.1
051910.KS	1	-14899	-29787.9	-29757.6
	2	-15105.9	-30189.8	-30122.9
	3	-15120	-30205.9	-30102.6
	4	-15119.7	-30193.5	-30053.7
	5	-15122.2	-30186.4	-30010.2
005490.KS	1	-14835.7	-29661.3	-29630.9
	2	-15106	-30190.1	-30123.2
	3	-15110	-30185.9	-30082.6
	4	-15119.7	-30193.5	-30053.7
	5	-15118.8	-30179.5	-30003.2
006400.KS	1	-14386	-28761.9	-28731.5
	2	-14629.9	-29237.9	-29171
	3	-14662.1	-29290.3	-29187
	4	-14663.7	-29281.5	-29141.7
	5	-14672	-29285.9	-29109.6
009150.KS	1	-14923.7	-29837.3	-29807
	2	-15098.5	-30175	-30108.1
	3	-15134	-30234	-30130.7
	4	-15139.9	-30233.8	-30094
	5	-15149.7	-30241.4	-30065.1
010140.KS	1	-14460.1	-28910.2	-28879.8
	2	-14717.3	-29412.6	-29345.8
	3	-14740.7	-29447.4	-29344
	4	-14739.6	-29433.1	-29293.3
	5	-14751	-29444	-29267.8
000880.KS	1	-14829.3	-29648.6	-29618.2
	2	-14947.1	-29872.2	-29805.3
	3	-14986.7	-29939.5	-29836.1
	4	-14994.1	-29942.1	-29802.3
	5	-15006.7	-29955.3	-29779
000720.KS	1	-14518.4	-29026.8	-28996.4
	2	-14705.9	-29389.8	-29322.9
	3	-14728.4	-29422.9	-29319.5
	4	-14734.4	-29422.7	-29282.9
	5	-14735.8	-29413.7	-29237.4
009540.KS	1	-13639.7	-27269.4	-27239
	2	-13974.4	-27926.8	-27859.9
	3	-14012.5	-27991	-27887.6
	4	-13974.5	-27903	-27763.2
	5	-14032.5	-28007	-27830.7

Table C.6: GMM Fitting: KOSPI50 30 minute data

Ticker	Number of Components	Negative Loglikelihood	AIC	BIC
034220.KS	1	-10589.6	-21169.3	-21140.1
	2	-10942.6	-21863.3	-21799.0
	3	-10948.5	-21863	-21763.7
	4	-10960.7	-21875.4	-21741
	5	-10966.3	-21874.6	-21705.2
066570.KS	1	-11328.4	-22646.9	-22617.7
	2	-11623.4	-23224.8	-23160.6
	3	-11629.2	-23224.4	-23125.1
	4	-11641.2	-23236.4	-23102
	5	-11645.9	-23233.9	-23064.5
051910.KS	1	-10853.7	-21697.4	-21668.2
	2	-11394.7	-22767.5	-22703.2
	3	-11395.2	-22756.4	-22657.1
	4	-11396.4	-22746.9	-22612.5
	5	-11427.9	-22797.8	-22628.3
005490.KS	1	-10798.4	-21586.7	-21557.5
	2	-11188.2	-22354.3	-22290.1
	3	-11195.3	-22356.5	-22257.2
	4	-11194.6	-22343.3	-22208.9
	5	-11196.5	-22334.9	-22165.5
006400.KS	1	-10206.6	-20403.3	-20374.1
	2	-10632.7	-21243.4	-21179.2
	3	-10638.5	-21243	-21143.7
	4	-10642.3	-21238.6	-21104.3
	5	-10640.3	-21222.6	-21053.2
009150.KS	1	-10903.3	-21796.7	-21767.5
	2	-11320	-22617.9	-22553.7
	3	-11321.2	-22608.4	-22509.1
	4	-11336.8	-22627.6	-22493.3
	5	-11326.9	-22595.8	-22426.4
010140.KS	1	-10243.6	-20477.3	-20448.1
	2	-10823.1	-21624.3	-21560.0
	3	-10826.8	-21619.5	-21520.2
	4	-10840.2	-21634.3	-21499.9
	5	-10854.1	-21650.2	-21480.8
000880.KS	1	-10864.4	-21718.8	-21689.6
	2	-11057.7	-22093.5	-22029.3
	3	-11063.3	-22092.6	-21993.3
	4	-11063.3	-22080.6	-21946.3
	5	-11066.7	-22075.4	-21906.1
000720.KS	1	-10629.2	-21248.3	-21219.1
	2	-10976.1	-21930.1	-21865.9
	3	-10992.1	-21950.2	-21850.9
	4	-10993	-21939.9	-21805.5
	5	-10998.8	-21939.7	-21770.2
009540.KS	1	-10006.1	-20002.2	-19973
	2	-10326.6	-20631.2	-20566.9
	3	-10326.6	-20619.2	-20519.9
	4	-10337.6	-20629.2	-20494.8
	5	-10344.2	-20630.5	-20461.1

APPENDIX D

SHARPE RATIO FOR ALL TRANSACTION COST ASSUMPTION

Table D.1: Sharpe ratio of trading of the Discrete Hidden Markov Model in SET50 (assume no transaction cost)

	5 minutes			10 minutes			30 minutes			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	132.8862%	144.1766%	166.8060%	44.4074%	14.5011%	33.1477%	27.9917%	5.5993%	7.3490%	2.0124%
σ	19.3749%	17.2391%	15.3699%	3.9482%	9.5107%	26.0692%	5.4309%	4.5668%	3.2137%	1.1801%
Sharpe Ratio	6.8536	8.3576	10.8463	11.2223	1.5143	1.2677	5.1359	1.2044	2.2559	1.6213

Table D.2: Sharpe ratio of trading of the Discrete Hidden Markov Model in SET50 (assume 0.05% transaction cost)

	5 minutes			10 minutes			30 minutes			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	62.4152%	76.5519%	99.0951%	23.3392%	3.2343%	10.7075%	16.2559%	0.8825%	1.7823%	2.0124%
σ	26.4160%	15.8375%	17.9793%	4.8111%	8.5142%	22.0272%	5.6533%	3.6992%	3.5969%	1.1801%
Sharpe Ratio	2.3590	4.8273	5.5061	4.8305	0.3682	0.4816	2.8579	0.2118	0.4679	1.6213

Table D.3: Sharpe ratio of trading of the Discrete Hidden Markov Model in SET50 (assume 0.1% transaction cost)

	5 minutes			10 minutes			30 minutes			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	-8.0552%	8.9283%	31.3848%	2.2713%	-8.0324%	-11.7324%	4.5203%	-3.8342%	-3.7844%	2.0124%
σ	33.6317%	17.4080%	20.6029%	6.2650%	8.7144%	18.2320%	6.5836%	3.3489%	4.0199%	1.1801%
Sharpe Ratio	-0.2425	0.5072	1.5185	0.3467	-0.9331	-0.6489	0.6715	-1.1745	-0.9661	1.6213

Table D.4: Sharpe ratio of trading of the Continuous Hidden Markov Model in SET50 (assume no transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	0.9963%	1.0070%	8.6719%	3.8141%	3.9910%	0.5971%	3.3587%	1.2475%	5.0152%	0.9963%
σ	3.4152%	2.6644%	1.7642%	5.9260%	4.8509%	3.4087%	4.0215%	1.9937%	3.6649%	1.1801%
Sharpe Ratio	0.2627	0.3407	4.8592	0.6269	0.8023	0.1461	0.8105	0.5760	1.3414	0.2627

Table D.5: Sharpe ratio of trading of the Continuous Hidden Markov Model in SET50 (assume 0.05% transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	0.1361%	-0.0948%	7.3997%	3.0263%	3.4718%	-0.1371%	3.2538%	1.1178%	4.8025%	2.0124%
σ	3.6048%	2.8341%	1.6411%	5.8075%	5.3254%	3.7583%	3.9488%	1.9440%	3.6249%	1.1801%
Sharpe Ratio	0.0102	-0.0684	4.4485	0.5040	0.6333	-0.0629	0.7989	0.5240	1.2975	1.6213

Table D.6: Sharpe ratio of trading of the Continuous Hidden Markov Model in SET50 (assume 0.1% transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	-0.7238%	-1.1960%	6.1281%	2.2390%	2.9529%	-0.8708%	3.1490%	0.9883%	4.5900%	2.0124%
σ	3.8350%	3.0386%	1.6431%	5.7139%	5.8001%	4.1078%	3.8763%	1.8944%	3.5907%	1.1801%
Sharpe Ratio	-0.2146	-0.4263	3.6692	0.3745	0.4920	-0.2361	0.7868	0.4694	1.2507	1.6213

Table D.7: Table: Sharpe ratio of trading of the Discrete Hidden Markov Model in KOSPI50 (assume no transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	6.0284%	8.5844%	24.7226%	9.4293%	3.2597%	2.3606%	-0.0012%	1.6554%	-0.0616%	1.4687%
σ	8.8448%	8.2512%	8.1236%	4.2540%	5.8098%	2.3981%	0.0020%	1.7895%	2.4847%	2.1615%
Sharpe Ratio	0.6676	1.0254	3.0281	2.1875	0.5397	0.9327	-61.0675	0.8559	-0.0746	0.6221

Table D.8: Sharpe ratio of trading of the Discrete Hidden Markov Model in KOSPI50 (assume 0.05% transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	-1.5605%	1.2178%	1.0058%	-2.1209%	-5.6903%	-3.4060%	-0.9012%	-1.7613%	-3.4782%	1.4687%
σ	3.3043%	1.0524%	8.6708%	2.8765%	4.0859%	2.7977%	1.5609%	1.0726%	3.1533%	2.1615%
Sharpe Ratio	-0.5098	1.0394	0.1017	-0.7804	-1.4230	-1.2617	-0.6567	-1.7575	-1.1423	0.6221

Table D.9: Sharpe ratio of trading of the Discrete Hidden Markov Model in KOSPI50 (assume 0.1% transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	-9.1495%	-6.1489%	-22.7109%	-13.6710%	-14.6403%	-9.1727%	-1.8012%	-5.1780%	-6.8949%	1.4687%
σ	13.7483%	7.8752%	9.8700%	6.0541%	9.1296%	5.8485%	3.1197%	0.9542%	3.8360%	2.1615%
Sharpe Ratio	-0.6745	-0.7965	-2.3136	-2.2786	-1.6172	-1.5896	-0.6171	-5.5565	-1.8297	0.6221

Table D.10: Sharpe ratio of trading of the Continuous Hidden Markov Model in KOSPI50 (assume no transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	7.6214%	2.1746%	2.7673%	0.6668%	1.2595%	3.4084%	2.9220%	1.8632%	4.0636%	1.4687%
σ	5.9083%	5.9179%	1.4997%	8.5185%	5.5679%	6.0435%	10.6972%	11.6194%	11.4362%	2.1615%
Sharpe Ratio	1.2732	0.3507	1.7791	0.0666	0.2084	0.5476	0.2639	0.1518	0.3467	0.6221

Table D.11: Sharpe ratio of trading of the Continuous Hidden Markov Model in KOSPI50 (assume 0.05% transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	6.7943%	1.1443%	1.6548%	-0.0611%	0.4325%	2.6582%	2.5741%	1.4479%	3.5543%	1.4687%
σ	6.2202%	6.2382%	1.8425%	9.1623%	5.8634%	6.4040%	10.8685%	11.8982%	11.6520%	2.1615%
Sharpe Ratio	1.0763	0.1675	0.8443	-0.0175	0.0568	0.3996	0.2277	0.1134	0.2965	0.6221

Table D.12: Sharpe ratio of trading of the Continuous Hidden Markov Model in KOSPI50 (assume 0.1% transaction cost)

	5 minutes			**10 minutes**			**30 minutes**			
	3 state	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**3 states**	**4 states**	**5 states**	**Benchmark**
μ	5.9674%	0.1146%	0.5425%	-0.7885%	-0.3942%	1.9084%	2.2266%	1.0329%	3.0455%	1.4687%
σ	6.5321%	6.5593%	2.1884%	9.8074%	6.1594%	6.7691%	11.0426%	12.1803%	11.8700%	2.1615%
Sharpe Ratio	0.8984	0.0024	0.2026	-0.0905	-0.0801	0.2673	0.1927	0.0767	0.2482	0.6221

APPENDIX E
JENSON'S ALPHA: REGRESSION ANALYSIS

Table E.1: Jenson's Alpha (Discrete case, SET50, no transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	0.06297	0.00572	11.00363	0.00000
		slope	3.93881	1.87369	2.10217	0.03974
	4 states	Intercept	0.06850	0.00573	11.95991	0.00000
		slope	3.72206	2.04160	1.82311	0.07327
	5 states	Intercept	0.08022	0.00432	18.56216	0.00000
		slope	1.32523	1.53372	0.86406	0.39099
10 min	3 states	Intercept	0.02092	0.00275	7.61942	0.00000
		slope	1.55349	1.03880	1.49546	0.14004
	4 states	Intercept	0.00692	0.00190	3.63945	0.00057
		slope	0.10959	0.54414	0.20140	0.84107
	5 states	Intercept	0.01502	0.00391	3.84572	0.00029
		slope	2.95917	1.36943	2.16087	0.03471
30 min	3 states	Intercept	0.01339	0.00157	8.53016	0.00000
		slope	0.28046	0.69591	0.40301	0.68837
	4 states	Intercept	0.00283	0.00134	2.11024	0.03902
		slope	-0.57553	0.82434	-0.69817	0.48777
	5 states	Intercept	0.00329	0.00133	2.47039	0.01636
		slope	0.63770	0.37673	1.69271	0.09570

Table E.2: Jenson's Alpha (Discrete case, SET50, 0.05% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	0.02872	0.00538	5.33643	0.00000
		slope	4.40587	1.80102	2.44632	0.01738
	4 states	Intercept	0.03571	0.00494	7.23223	0.00000
		slope	3.93404	1.78874	2.19934	0.03172
	5 states	Intercept	0.04734	0.00399	11.86487	0.00000
		slope	1.68980	1.43897	1.17431	0.24491
10 min	3 states	Intercept	0.01056	0.00233	4.52421	0.00003
		slope	2.07871	0.91569	2.27009	0.02681
	4 states	Intercept	0.00145	0.00175	0.82488	0.41271
		slope	0.17092	0.43919	0.38916	0.69853
	5 states	Intercept	0.00409	0.00344	1.19164	0.23809
		slope	3.18249	1.40427	2.26630	0.02705
30 min	3 states	Intercept	0.00770	0.00157	4.89584	0.00001
		slope	0.33346	0.71038	0.46941	0.64048
	4 states	Intercept	0.00050	0.00137	0.36742	0.71460
		slope	-0.43771	0.78439	-0.55802	0.57890
	5 states	Intercept	0.00055	0.00134	0.41327	0.68088
		slope	0.76181	0.41203	1.84892	0.06940

Table E.3: Jenson's Alpha (Discrete case, SET50, 0.1% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	-0.00553	0.00528	-1.04680	0.29939
		slope	4.87291	1.77931	2.73866	0.00811
	4 states	Intercept	0.00292	0.00432	0.67683	0.50112
		slope	4.14599	1.56298	2.65261	0.01021
	5 states	Intercept	0.01446	0.00392	3.68996	0.00049
		slope	2.05434	1.39421	1.47348	0.14585
10 min	3 states	Intercept	0.00020	0.00203	0.09778	0.92243
		slope	2.60393	0.81802	3.18323	0.00231
	4 states	Intercept	-0.00402	0.00192	-2.09497	0.04040
		slope	0.23225	0.51860	0.44785	0.65588
	5 states	Intercept	-0.00684	0.00315	-2.16988	0.03399
		slope	3.40580	1.46250	2.32875	0.02326
30 min	3 states	Intercept	0.00200	0.00163	1.22730	0.22451
		slope	0.38647	0.73207	0.52791	0.59951
	4 states	Intercept	-0.00182	0.00146	-1.24947	0.21635
		slope	-0.29988	0.75399	-0.39773	0.69224
	5 states	Intercept	-0.00218	0.00143	-1.52322	0.13296
		slope	0.88593	0.46106	1.92152	0.05942

Table E.4: Jenson's Alpha (Continuous case, SET50, no transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	-0.00034	0.00096	-0.35220	0.72593
		slope	2.35134	0.46956	5.00754	0.00001
	4 states	Intercept	-0.00026	0.00136	-0.19270	0.84784
		slope	2.13434	0.52667	4.05252	0.00015
	5 states	Intercept	0.00331	0.00165	2.00648	0.04932
		slope	2.55293	0.85468	2.98699	0.00408
10 min	3 states	Intercept	0.00090	0.00200	0.45018	0.65420
		slope	2.74933	0.94798	2.90021	0.00520
	4 states	Intercept	0.00100	0.00114	0.87754	0.38369
		slope	2.68676	0.68117	3.94435	0.00021
	5 states	Intercept	-0.00056	0.00117	-0.47380	0.63736
		slope	2.43096	0.72562	3.35019	0.00140
30 min	3 states	Intercept	0.00074	0.00167	0.44020	0.66138
		slope	2.56615	0.55938	4.58748	0.00002
	4 states	Intercept	-0.00021	0.00148	-0.14008	0.88906
		slope	2.32716	0.57412	4.05343	0.00015
	5 states	Intercept	0.00148	0.00115	1.27796	0.20619
		slope	2.76291	0.72796	3.79540	0.00035

Table E.5: Jenson's Alpha (Continuous case, SET50, 0.05% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	-0.00076	0.00155	-0.48735	0.62779
		slope	2.35911	0.65034	3.62749	0.00059
	4 states	Intercept	-0.00081	0.00174	-0.46607	0.64286
		slope	2.18861	0.73122	2.99309	0.00401
	5 states	Intercept	0.00269	0.00213	1.26325	0.21139
		slope	2.56968	0.89342	2.87624	0.00556
10 min	3 states	Intercept	0.00052	0.00226	0.22982	0.81902
		slope	2.74045	0.94916	2.88723	0.00540
	4 states	Intercept	0.00076	0.00199	0.38166	0.70406
		slope	2.66650	0.83448	3.19541	0.00223
	5 states	Intercept	-0.00091	0.00221	-0.41270	0.68130
		slope	2.42797	0.92477	2.62547	0.01096
30 min	3 states	Intercept	0.00068	0.00191	0.35728	0.72214
		slope	2.58314	0.79908	3.23264	0.00199
	4 states	Intercept	-0.00027	0.00188	-0.14424	0.88580
		slope	2.32965	0.78866	2.95394	0.00448
	5 states	Intercept	0.00137	0.00183	0.74937	0.45656
		slope	2.76312	0.76797	3.59796	0.00065

Table E.6: Jenson's Alpha (Continuous case, SET50, 0.1% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	-0.00117	0.00102	-1.14741	0.25577
		slope	2.36687	0.48963	4.83395	0.00001
	4 states	Intercept	-0.00136	0.00139	-0.98382	0.32915
		slope	2.24290	0.53360	4.20333	0.00009
	5 states	Intercept	0.00207	0.00172	1.20545	0.23276
		slope	2.58640	0.87588	2.95293	0.00449
10 min	3 states	Intercept	0.00014	0.00202	0.07035	0.94415
		slope	2.73156	0.95042	2.87406	0.00560
	4 states	Intercept	0.00051	0.00119	0.43288	0.66665
		slope	2.64620	0.68130	3.88402	0.00026
	5 states	Intercept	-0.00126	0.00124	-1.01856	0.31250
		slope	2.42495	0.74629	3.24935	0.00190
30 min	3 states	Intercept	0.00062	0.00167	0.37347	0.71011
		slope	2.60012	0.55853	4.65530	0.00002
	4 states	Intercept	-0.00033	0.00148	-0.22618	0.82183
		slope	2.33214	0.57437	4.06033	0.00014
	5 states	Intercept	0.00127	0.00113	1.11866	0.26775
		slope	2.76332	0.72936	3.78871	0.00035

Table E.7: Jenson's Alpha (Discrete case, KOSPI50, no transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	0.00280	0.00143	1.95021	0.05575
		slope	0.03391	0.09676	0.35043	0.72722
	4 states	Intercept	0.00405	0.00209	1.94088	0.05690
		slope	-0.05419	0.13150	-0.41214	0.68168
	5 states	Intercept	0.01159	0.00241	4.80715	0.00001
		slope	0.30420	0.22035	1.38050	0.17247
10 min	3 states	Intercept	0.00451	0.00121	3.72957	0.00042
		slope	-0.20204	0.24669	-0.81899	0.41598
	4 states	Intercept	0.00148	0.00085	1.74941	0.08525
		slope	0.02440	0.08387	0.29093	0.77209
	5 states	Intercept	0.00109	0.00053	2.04148	0.04554
		slope	-0.05149	0.05436	-0.94736	0.34719
30 min	3 states	Intercept	-0.00002	0.00022	-0.09120	0.92764
		slope	-0.09781	0.07319	-1.33645	0.18637
	4 states	Intercept	0.00080	0.00070	1.14691	0.25590
		slope	-0.17337	0.14592	-1.18816	0.23938
	5 states	Intercept	-0.00012	0.00082	-0.15002	0.88124
		slope	0.08855	0.13689	0.64685	0.52016

Table E.8: Jenson's Alpha (Discrete case, KOSPI50, 0.05% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	-0.00085	0.00062	-1.36643	0.17682
		slope	0.12419	0.11424	1.08710	0.28127
	4 states	Intercept	0.00045	0.00083	0.54786	0.58579
		Slope	0.17233	0.08384	2.05555	0.04411
	5 states	Intercept	0.00028	0.00155	0.18078	0.85714
		Slope	0.35100	0.19263	1.82211	0.07334
10 min	3 states	Intercept	-0.00099	0.00102	-0.97218	0.33480
		Slope	-0.20259	0.17676	-1.14613	0.25622
	4 states	Intercept	-0.00276	0.00090	-3.07415	0.00316
		Slope	-0.01946	0.09214	-0.21116	0.83347
	5 states	Intercept	-0.00166	0.00078	-2.14127	0.03626
		slope	-0.04272	0.05371	-0.79545	0.42944
30 min	3 states	Intercept	-0.00042	0.00036	-1.16637	0.24800
		slope	-0.17758	0.10339	-1.71762	0.09094
	4 states	Intercept	-0.00083	0.00084	-0.99094	0.32563
		slope	-0.16460	0.14618	-1.12596	0.26459
	5 states	Intercept	-0.00172	0.00095	-1.81365	0.07465
		slope	0.02500	0.12511	0.19981	0.84230

Table E.9: Jenson's Alpha (Discrete case, KOSPI50, 0.1% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	-0.00450	0.00195	-2.31384	0.02406
		slope	0.21448	0.26047	0.82343	0.41347
	4 states	Intercept	-0.00315	0.00135	-2.33355	0.02293
		slope	0.39887	0.18775	2.12446	0.03769
	5 states	Intercept	-0.01103	0.00201	-5.50088	0.00000
		slope	0.39780	0.22459	1.77125	0.08151
10 min	3 states	Intercept	-0.00649	0.00161	-4.02326	0.00016
		slope	-0.20313	0.15887	-1.27862	0.20587
	4 states	Intercept	-0.00700	0.00162	-4.31427	0.00006
		slope	-0.06331	0.13125	-0.48236	0.63128
	5 states	Intercept	-0.00441	0.00136	-3.24435	0.00191
		slope	-0.03395	0.09102	-0.37299	0.71045
30 min	3 states	Intercept	-0.00081	0.00058	-1.40089	0.16631
		slope	-0.25734	0.13699	-1.87848	0.06510
	4 states	Intercept	-0.00246	0.00106	-2.32540	0.02339
		slope	-0.15582	0.15280	-1.01971	0.31189
	5 states	Intercept	-0.00333	0.00113	-2.93116	0.00475
		slope	-0.03855	0.11773	-0.32748	0.74443

Table E.10: Jenson's Alpha (Continuous case, KOSPI50, no transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	0.00338	0.00205	1.65201	0.10367
		slope	0.46430	0.22131	2.09796	0.04006
	4 states	Intercept	0.00087	0.00190	0.45868	0.64810
		slope	0.26279	0.28960	0.90743	0.36775
	5 states	Intercept	0.00100	0.00145	0.68961	0.49306
		slope	0.63756	0.26628	2.39429	0.01974
10 min	3 states	Intercept	0.00013	0.00195	0.06626	0.94738
		slope	0.32293	0.30515	1.05828	0.29410
	4 states	Intercept	0.00046	0.00197	0.23437	0.81548
		slope	0.19831	0.22828	0.86872	0.38841
	5 states	Intercept	0.00141	0.00229	0.61740	0.53927
		slope	0.37471	0.26544	1.41163	0.16314
30 min	3 states	Intercept	0.00132	0.00208	0.63604	0.52713
		slope	0.02431	0.21763	0.11172	0.91141
	4 states	Intercept	0.00085	0.00235	0.36143	0.71902
		slope	-0.05478	0.18308	-0.29922	0.76579
	5 states	Intercept	0.00192	0.00245	0.78147	0.43755
		slope	-0.09722	0.18947	-0.51314	0.60971

Table E.11: Jenson's Alpha (Continuous case, KOSPI50, 0.05% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	0.00299	0.00205	1.45514	0.15076
		slope	0.46458	0.22021	2.10973	0.03899
	4 states	Intercept	0.00038	0.00191	0.19750	0.84409
		slope	0.27108	0.28493	0.95138	0.34517
	5 states	Intercept	0.00048	0.00147	0.32415	0.74693
		slope	0.63291	0.27200	2.32686	0.02331
10 min	3 states	Intercept	-0.00022	0.00197	-0.11374	0.90982
		slope	0.34025	0.30512	1.11514	0.26916
	4 states	Intercept	0.00007	0.00197	0.03491	0.97227
		slope	0.19578	0.22615	0.86571	0.39004
	5 states	Intercept	0.00105	0.00230	0.45817	0.64846
		slope	0.37937	0.26423	1.43577	0.15618
30 min	3 states	Intercept	0.00116	0.00208	0.55657	0.57986
		slope	0.01775	0.21610	0.08216	0.93479
	4 states	Intercept	0.00065	0.00237	0.27600	0.78348
		slope	-0.05843	0.18161	-0.32174	0.74875
	5 states	Intercept	0.00168	0.00245	0.68566	0.49552
		slope	-0.11639	0.19171	-0.60712	0.54602

Table E.12: Jenson’s Alpha (Continuous case, KOSPI50, 0.1% transaction cost)

Frequency	States		Coefficient	SE	t-stat	p-value
5 min	3 states	Intercept	0.00260	0.00206	1.25738	0.21341
		slope	0.46486	0.21924	2.12032	0.03805
	4 states	Intercept	-0.00012	0.00192	-0.06035	0.95207
		slope	0.27936	0.28052	0.99586	0.32325
	5 states	Intercept	-0.00005	0.00148	-0.03548	0.97182
		slope	0.62826	0.27855	2.25542	0.02771
10 min	3 states	Intercept	-0.00058	0.00200	-0.28945	0.77322
		slope	0.35757	0.30516	1.17175	0.24585
	4 states	Intercept	-0.00032	0.00197	-0.16410	0.87020
		slope	0.19325	0.22424	0.86180	0.39217
	5 states	Intercept	0.00070	0.00232	0.30044	0.76487
		slope	0.38402	0.26324	1.45883	0.14974
30 min	3 states	Intercept	0.00100	0.00209	0.47725	0.63489
		slope	0.01120	0.21469	0.05215	0.95858
	4 states	Intercept	0.00046	0.00239	0.19181	0.84853
		slope	-0.06208	0.18030	-0.34434	0.73177
	5 states	Intercept	0.00145	0.00245	0.58957	0.55766
		slope	-0.13556	0.19447	-0.69709	0.48840

APPENDIX F

SIGNAL COUNT

Table F.1: Total number of signals that resulted in positive or negative price movement (Discrete case, SET50)

Frequency	**# of states**	ADVANC.BK	BANPU.BK	BCP.BK	CPF.BK	DTAC.BK	IRPC.BK	IVL.BK	PTTEP.BK	TCAP.BK	TRUE.BK
5 min	3	194	130	101	242	162	103	197	251	155	150
	4	226	88	69	269	161	107	203	180	170	241
	5	161	114	104	286	151	148	237	82	156	193
10 min	3	18	19	71	41	33	70	37	24	49	74
	4	15	26	12	42	30	2	22	0	0	33
	5	56	47	42	79	64	34	53	25	30	27
30 min	3	40	28	40	36	18	37	51	2	62	33
	4	5	16	10	10	5	14	29	8	0	6
	5	10	13	5	6	6	8	16	7	9	26

Table F.2: Total number of signals that resulted in positive price movement (Discrete case, SET50)

Frequency	**# of states**	ADVANC.BK	BANPU.BK	BCP.BK	CPF.BK	DTAC.BK	IRPC.BK	IVL.BK	PTTEP.BK	TCAP.BK	TRUE.BK
5 min	3	160	87	82	157	134	81	154	186	139	131
	4	175	66	62	216	135	89	151	129	147	202
	5	129	92	93	244	128	132	191	58	138	163
10 min	3	13	14	65	35	19	51	26	13	35	54
	4	13	14	8	29	20	2	14	-	-	24
	5	41	21	36	61	47	29	33	10	20	17
30 min	3	24	19	35	25	13	27	40	1	52	25
	4	2	4	10	4	4	8	19	6	-	5
	5	7	7	3	3	5	5	11	5	4	15

Table F.3: Total number of signals that resulted in positive or negative price movement (Continuous case, SET50)

Frequency	**# of states**	ADVANC.BK	BANPU.BK	BCP.BK	CPF.BK	DTAC.BK	IRPC.BK	IVL.BK	PTTEP.BK	TCAP.BK	TRUE.BK
5 min	3	866	726	832	577	740	719	663	1198	922	915
	4	725	725	922	765	890	696	730	1042	933	783
	5	728	791	859	736	808	774	712	861	860	669
10 min	3	476	405	548	330	699	542	439	730	475	443
	4	468	495	652	336	580	552	475	774	360	431
	5	488	479	582	301	573	451	499	774	476	396
30 min	3	121	335	313	99	380	231	338	454	329	180
	4	114	324	284	113	390	251	343	450	302	164
	5	115	318	259	127	400	206	324	454	279	186

Table F.4: Total number of signals that resulted in positive price movement (Continuous case, SET50)

Frequency	# of states	ADVANC.BK	BANPU.BK	BCP.BK	CPF.BK	DTAC.BK	IRPC.BK	IVL.BK	PTTEP.BK	TCAP.BK	TRUE.BK
5 min	3	437	356	415	286	376	362	336	600	469	447
	4	369	349	462	376	445	348	374	523	474	380
	5	368	382	428	365	405	387	360	438	436	323
10 min	3	246	193	276	157	346	271	223	355	238	217
	4	239	237	325	167	284	273	236	381	178	207
	5	246	228	293	150	278	225	249	377	241	194
30 min	3	62	159	158	42	202	121	174	221	172	84
	4	59	156	141	55	206	128	172	221	161	76
	5	57	148	131	57	210	105	167	224	147	89

Table F.5: Total number of signals that resulted in positive or negative price movement (Discrete case, KOSPI50)

Frequency	# of states	034220	066570	051910	005490	006400	009150	010140	000880	000720	009540
5 min	3	0	0	32	84	46	0	0	0	0	0
	4	0	0	39	55	3	19	0	0	0	43
	5	63	26	72	70	44	81	3	0	8	180
10 min	3	12	33	58	84	40	0	2	12	0	45
	4	10	29	89	42	15	4	2	8	0	7
	5	12	16	61	5	10	2	5	0	14	11
30 min	3	0	0	0	20	0	0	0	0	0	0
	4	0	1	21	39	6	4	0	0	2	23
	5	0	4	9	30	21	6	3	7	1	16

Table F.6: Total number of signals that resulted in positive price movement (Discrete case, KOSPI50)

Frequency	# of states	034220	066570	051910	005490	006400	009150	010140	000880	000720	009540
5 min	3	0	0	25	61	31	0	0	0	0	0
	4	0	0	26	44	0	10	0	0	0	41
	5	34	20	58	54	32	50	2	0	6	143
10 min	3	7	14	41	50	30	0	2	6	0	33
	4	2	18	51	24	12	4	2	4	0	7
	5	10	8	34	2	5	2	2	0	6	8
30 min	3	0	0	0	12	0	0	0	0	0	0
	4	0	0	15	19	5	2	0	0	0	11
	5	0	1	6	15	14	5	1	3	0	6

Table F.7: Total number of signals that resulted in positive or negative price movement (Continuous case, KOSPI)

Frequency	# of states	034220	066570	051910	005490	006400	009150	010140	000880	000720	009540
5 min	3	1761	1336	1513	1959	1716	1635	945	1537	1348	1574
	4	1435	1442	1538	1893	1856	1646	1143	1727	1325	1462
	5	1499	1286	1716	2066	1496	1540	961	1730	1418	1313
10 min	3	870	976	894	1274	1006	884	678	956	630	635
	4	833	826	912	1068	884	791	512	797	729	725
	5	962	935	902	1191	983	793	507	895	846	895
30 min	3	368	297	437	500	313	339	375	272	337	292
	4	414	326	425	440	333	354	354	377	301	301
	5	371	297	358	446	343	347	371	371	276	300

Table F.8: Total number of signals that resulted in positive price movement (Continuous case, KOSPI50)

Frequency	# of states	034220	066570	051910	005490	006400	009150	010140	000880	000720	009540
5 min	3	877	687	782	982	873	814	450	750	678	774
	4	721	745	778	943	944	815	555	845	683	721
	5	753	664	870	1035	763	762	485	840	702	640
10 min	3	432	502	448	627	517	443	334	464	313	312
	4	414	430	462	529	455	395	247	382	346	356
	5	472	486	458	591	507	376	248	442	409	439
30 min	3	162	159	219	244	171	163	172	135	168	140
	4	194	174	212	210	180	179	169	176	149	137
	5	170	160	173	214	186	174	171	173	133	136

BIOGRAPHY

Name:	Mr. Po-Lin Wu
Date of Birth:	16^{th} May 1990
Educational Attainment:	Academic Year: 2008
	Bachelor of Science
	Mahidol University International College
	Academic Year: 2015
	Master in Finance
	Thammasat Business School
Work Experiences	Senior Officer
	Tycoons Worldwide Group(Thailand) PCL