การอนุมานด้วยความน่าจะเป็นแบบเบย์สำหรับพลวัตรของราคาหุ้น ในระดับบรรษัท

นางสาวไฮเซน วาง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2559

BAYESIAN PROBABILISTIC INFERENCE ON FIRM-LEVEL STOCK PRICE DYNAMICS

Haizhen Wang

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Applied Mathematics

Suranaree University of Technology

Academic Year 2016

BAYESIAN PROBABILISTIC INFERENCE ON FIRM-LEVEL STOCK PRICE DYNAMICS

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

(Asst. Prof. Dr. Eckart Schulz)
Chairperson

P. Sattayatham
(Prof. Dr. Pairote Sattayatham)
Member (Thesis Advisor)

(Prof. Dr. Bhusana Premanode)
Member

(Asst. Prof. Dr. Arjuna Chaiyasena)
Member

(Asst. Prof. Dr. Benjawan Rodjanadid)
Member

(Prof. Dr. Sukit Limpijumnong)
Vice Rector for Academic Affairs and Innovation

(Prof. Dr. Santi Maensiri)
Dean of Institute of Science

ไฮเซน วาง : การอนุมานด้วยความน่าจะเป็นแบบเบย์สำหรับพลวัตรของราคาหุ้นในระดับบรรษัท(BAYESIAN PROBABILISTIC INFERENCE ON FIRM-LEVEL STOCK PRICE DYNAMICS) อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.ไพโรจน์ สัตยธรรม, 104 หน้า.

การอนุมานแบบเบย์/เครือข่ายพลวัตแบบเบย์/การลงทุนขั้นพื้นฐาน/สัดส่วนพีอี/รางวัลความเสี่ยงในส่วนของผู้ถือหุ้น

เมื่อพิจารณาถึงการตัดสินใจเกี่ยวกับการลงทุนในตลาดหลักทรัพย์แต่ละวัน จากมุมมองเชิงวิชาการในเศรษฐมิติเชิงการเงิน และการเรียนรู้ด้วยเครื่อง แทบจะไม่พิจารณาการประมาณค่าราคาหุ้นในเชิงพื้นฐานเลย แต่ในวิทยานิพนธ์ฉบับนี้เราเสนอการประมาณค่าราคาหุ้นระดับบรรษัทที่นำข้อมูลเชิงพื้นฐานเข้ามาพิจารณาด้วย เมื่อพิจารณาถึงพฤติกรรมเชิงประจักษ์ของความผันผวน จะพบว่าราคาหุ้นจะเคลื่อนไปห่างไปจากมูลค่าหุ้นเป็นครั้งคราว เราจะสร้างสูตรสำหรับพลวัตของราคาหุ้นโดยการใช้พลวัตเครือข่ายเบย์ และใช้ขั้นตอนวิธีค่าคาดหวังสูงสุดในการคำนวณตัวแปรต่าง ๆ นอกจากนี้ยังใช้ขั้นตอนวิธีก้าวหน้า-ถอยหลังในการกรองและทำให้ราบเรียบ

ผลของการวิจัยได้มีการสร้างกลยุทธ์ที่ง่ายต่อการนำไปใช้ในการซื้อขายหุ้น เราได้นำกลยุทธ์นี้ไปทดลองซื้อขายหุ้นในตลาดต่างๆดังเช่นตลาด SET ของประเทศไทย ตลาด NYSE และ NASDAQ ของประเทศอเมริกา ด้วยการทดลองทำการซื้อขายหุ้นด้วยระยะเวลานานพอควรพบว่ากลยุทธ์ของเราให้ผลตอบแทนดีกว่ากลยุทธ์ซื้อ-และ-ถืออย่างมีนัยสำคัญ

สาขาวิชาคณิตศาสตร์ ลายมือชื่อนักศึกษา Haizhen Wang

ปีการศึกษา 2559 ลายมือชื่ออาจารย์ที่ปรึกษา P. Sattayatham

HAIZHEN WANG : BAYESIAN PROBABILISTIC INFERENCE ON FIRM-LEVEL STOCK PRICE DYNAMICS. THESIS ADVISOR : PROF. PAIROTE SATTAYATHAM, Ph.D. 104 PP.

BAYESIAN INFERENCE/ DYNAMIC BAYESIAN NETWORK/ FUNDAMENTAL INVESTMENT/ PE RATIO / EQUITY RISK PREMIUM

When considering daily investment decisions in a security market, recent academic developments in financial econometrics and machine learning have rarely looked at fundamental estimation. In this thesis, we take such fundamental information into account to estimate the firm-level stock price dynamics. Due to behavioral finance evidence of volatility, the stock price may temporarily sway away from its value from time to time. We simplify and formalize the stock price dynamics by employing an advanced dynamic Bayesian network (DBN) combined with the expectation maximization (EM) algorithm for calculating parameters and the forward-backward algorithm for filtering and smoothing.

A simple but practical trading strategy is invented based on the results of our model. We make stock trading experiments in the markets from different countries, namely from the SET in Thailand, NYSE and NASDAQ in US. Extensive experiments show that our trading strategy outperforms the buy-and-hold strategy significantly.

School of Mathematics | Student's Signature Haizhen Wang

Academic Year 2016 | Advisor's Signature P. Sattayatham

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Dr. Pairote Sattayatham, and Dr. Ratthachat Chatpatanasiri for their guidance throughout my studies. Without their help, I would not have finished my work.

In addition, many thanks go to Asst. Prof. Dr. Eckart Schulz. I would like to thank Prof. Dr.Bhusana Premanode, Asst. Prof. Dr. Arjuna Chaiyasena, and Asst.Prof. Dr. Benjawan Rodjanadid as my thesis committees to give me suggestions.

Guizhou University of Finance and Economics and Suranaree University of Technology both gave me the opportunity to study here for my Ph.D. My scincere appreciation is extended to the two universities.

Finally, I would like to thank my friends for sharing information, providing help, etc.

By far most of my thanks goes to my parents who have supported me during my long period of education. I would like to appreciate all the people who have appeared in my life, for their love, support and harassment, to make me feel happy and become strong.

Haizhen Wang

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DBN	Dynamic Bayesian Network
BN	Bayesian Network
EM	Expectation Maximization
HMM	Hidden Markov Model
CI	Conditional Independence
pdf	probability distribution function
PE	Price Earnings
GGM	Gordon Growth Model
DDM	Dividend Discount Model
DGM	Directed Graphical Model
ML	Maximum Likelihood
MAP	Maximum a Posteriori
NYSE	New York Stock Exchange

CHAPTER I

INTRODUCTION

1.1 Introduction

With the rapid advancement of machine learning technology, recent works make attempts to incorporate these machine learning techniques to construct trading systems that support decisions of investors in security markets. For example, Yeh et al., 2011 developed a two-stage multiple-kernel learning algorithm applied to stock market forecasting problems. Lu et al., 2009 proposed a two-stage modeling approach using independent component analysis (ICA) and support vector regression to alleviate the influence of noise in financial time series. Wen et al., 2010 proposed an intelligent trading system in the stock market based on oscillation box prediction by combining stock box theory and the support vector machine (SVM) algorithm. Hassan, 2009 presented a combination of the hidden Markov model (HMM) and the fuzzy models for forecasting stock market data. Kao et al., 2013 proposed a stock price forecasting model which integrates wavelet transform, multivariate adaptive regression splines, and support vector regression to improve the forecast accuracy. Kazem et al., 2013 proposed a forecasting model based on chaotic mapping, firefly algorithm and support vector regression. These recent works have the following common philosophical theme.

Common Philosophy of Applying Machine Learning to Financial Data: There exist hidden patterns in financial time series. Complicated techniques (and their combinations) such as support vector machine (SVM), multiple kernel learning, independent component analysis (ICA), hidden Markov models (HMM), fuzzy modeling and so on, can help investors discover hidden patterns represented by complicated mathematical formulas. The retrieved formulas can be used as forecasting rules to predict the stocks' directional movements, which in turn can be incorporated into the investors' trading strategy (buy low and sell high, as predicted by the rules) to make an excess return in the market.

Based on the same philosophy mentioned above, existing works, however, have common limitations. Firstly, the discovered patterns are very complicated (highly non-linear) and lack financial interpretation. Note that, in general, higher degree of pattern complexities are more prone to make the training data over-fitting (Bishop, 2006). Secondly, each financial time-series has to be trained separately, resulting in one set of distinct patterns for each different security. In other words, there is no common pattern in the data of securities. Thirdly, because of pattern complexities, practical trading implementations are not easy for some investors. In fact, sophisticated trading programs have to be constructed by users themselves. Lastly, there is no direct way, which is consistent to the probabilistic framework, to incorporate existing expert information (such as professional security analysts recommendations) into the learning system. Fairly speaking, although having the mentioned limitations, the core philosophy of existing research matches the philosophy of one investor group, called technical analysts (J. Murphy, 1999; Shannon, 2008). Technical analysts believe in price patterns and do not pay much attention to economic interpretation of the patterns. Therefore, this line of existing research may benefit this group of investors.

On another side of investment practitioners, there is a group named fundamentalists whose trading strategies have clear financial interpretations and are based on well-defined financial information (Mark, 2011; Damodaran, 2012; Lynch and Rothchild, 2000). Price Earnings ratio (simply called PE ratio, to be defined below shortly) and the Gordon Growth model (GGM) (Campbell and Shiller, 1988) are two of the most widely applied valuation toolkits for fundamentalists to make their investment decisions (Damodaran, 2012; Henry et al., 2010). Also, investment recommendations by security analysts are often based on PE ratios (Carvell et al., 1989) or the Gordon Growth model. Nevertheless, it is unfortunate that recent academic advancements in financial econometrics and machine learning rarely look at these two tools. To our knowledge, there is currently no formal framework capable of integrating expert knowledge with historical financial time-series data.

In the thesis, we focus on applying a Bayesian statistical analysis to formalize the process of stock valuation. We apply the powerful framework of dynamic Bayesian network (DBN) (Bishop, 2006; K. P. Murphy, 2012) to model the valuation process. Among these techniques, SVM is a classifier known for its excellent discrimination performance in the binary decision problem without a model of time, whereas HMM owns excellent temporal modeling properties (Valstar and Pantic, 2007). In contrast to existing machine learning frameworks mentioned above on price pattern discovery where the discovered patterns have no meaning in finance, the interpretation of our model is well justified according to behavioral finance (Szyszka, 2013). In the thesis, we propose to apply the machine learning framework to formalize the valuation process which somehow rarely gets attention from academic researchers. Next, unlike existing works where there are different discovered patterns for different securities, our proposed trading strategy resulting from the Bayesian framework is unified, i.e. we propose a single trading strategy which can be applied to every security. The proposed strategy is simple and has a clear financial interpretation so that it can be easily applied by every practitioner.

Moreover, expert opinions can be naturally integrated into our Bayesian learning framework. Finally, as our proposed dynamic Bayesian network has non-standard structure compared to literatures (Bishop, 2006; K. P. Murphy, 2012), we have successfully derived a new inference formulas by applying the forward-backward methodology, and the new parameter estimation algorithm according to the concept of Expectation Maximization (EM) algorithm (Bishop, 2006; K. P. Murphy, 2012).

Note that in this thesis, we focus on investment in individual firm-level securities which are usually preferred by individual investors; in contrast to investment institutions whose investment strategy is usually on portfolio level based on Modern Portfolio Theory (Barber and Odean, 2011).

We make stock trading based on the Price Earning (PE) ratio and the Gordon Growth model (GGM), respectively. To avoid duplicated writing, here, we only give diagrams based on PE ratio. (i) We collect historical data and separate the data into the training data and the testing data as described in Section 4.2. (ii) Due to behavioral finance and mean reversion, we formalize the stock price dynamics by the DBN, see Section 3.2. (iii) In the DBN framework, conditional independent properties are used to simplify the probabilistic inference. Forward-backward algorithm and EM algorithm are used for the training data to estimate parameters as explained in Section 3.3. Here, we use Maximum a Posteriori (MAP) method which can be solved by EM algorithm, so that existing expert information is incorporated into the framework. (iv) Based on the estimated parameters, we use forward-backward algorithm again for the testing data and calculate the filtering formula and smoothing formula to estimate the stock dynamics. (v) Long-term strategy and medium-term strategy are invented based on the results of our model in Section 4.2. (vi) Finally, we compare those two strategies with the buy-and-hold strategy on individual firm-level, respectively, see Section 4.3. Furthermore, we do statistical significance in our experiments on portfolio level to make sure that our model is robust.

1.2 Outline of the thesis

The main purpose of the thesis is to formalize a process of stock valuation by employing an advanced dynamic Bayesian network (DBN) methodology.

The thesis is divided into 6 chapters and is organized as follows. Chapter II reviews the preliminaries such as fundamental valuation, mean reversion, Bayesian statistical methods, parameter estimation in which expectation maximization (EM) algorithm is described, and normality test for a distribution.

Chapter III explains the motivation based on Pricing Earning (PE) ratio and presents the modeling of stock pricing dynamics based on on behavioral finance and fundamental investment using PE Ratio. We simplify the model and derives the inference of our model combined with the expectation maximization (EM) algorithm for calculating parameters and the forward-backward algorithm for filtering and smoothing. Moreover, a simple but practical trading strategy is invented based on the result of our model.

Experiments of our model in Chapter III are described in Chapter IV. We make stock trading experiments in the markets from different countries, namely the NYSE and NASDAQ in US and the SET in Thailand. Extensive experiments show that our trading strategy on individual firm-level equipped with the inferred PE Ratio outperforms the buy-and-hold strategy. Furthermore, we do statistical significance in our experiments on portfolio level and our method beats the buy-and-hold strategy.

Chapter V presents to model stock pricing dynamics based on the Gordon

Growth model (GGM) to reformulate the stock price dynamics model to reflect an important variable in corporate finance and valuation, namely equity risk premium.

The conclusion is presented in Chapter VI.

CHAPTER II

PRELIMINARIES

We formalize the stock price dynamics combing fundamental value with Bayesian statistical methods. In this chapter, we give the preliminaries. Fundamental value theory is described in Section 2.1. Mean reversion is explained in Section 2.2. For Bayesian statistical methods, basic definitions are given and Bayesian Network is introduced in Section 2.3. In Bayesian statistical methods, we meet unknown parameters which can be calculated by expectation maximization (EM) algorithm, see Section 2.4.

2.1 Fundamental value

Definition 2.1. The fundamental value in any period is the present value of the expected dividend D_{t+1} and the future price P_{t+1} over the next period, discounted at the expected return r. P_t^* denotes the fundamental value of an asset at the end of period t (Brealey et al., 2012; Bodie, 2009).

$$P_t^* = \frac{P_{t+1} + D_{t+1}}{1+r} \tag{2.1}$$

If the fundamental value, or the investor's estimation for what the stock is really worth, exceeds the market price, then the stock is considered undervalued.

(Campbell and Shiller, 1988) suggested that the dividend yield has significant power for future returns and proved that theoretically, the dividend process and stock price are cointegrated. The fundamental value is also calculated by the Dividend Discount model (DDM) (Brealey et al., 2012; Bodie, 2009).

Theorem 2.1. *Dividend Discount model (DDM formula) The fundamental value of an asset is equal the present value of all expected future dividends, provided that the discount rate is constant and stock price is equal to its fundamental value.*

$$P_0^* = \frac{D_1}{(1+r)} + \frac{D_2}{(1+r)^2} + \frac{D_3}{(1+r)^3} + ... = \sum_{t=1}^{\infty} \frac{D_t}{(1+r)^t} \tag{2.2}$$

where P_0^* is the fundamental value of the asset, D_t is the expected dividend in the tth year, $t = 1, 2, ...,$ and r is the discount rate. The formula is called the dividend discount model (DDM) formula of the stock prices .

Corollary 2.2. *Gordon Growth model (GGM) If the dividends are trending upward at a constant growth rate g , then the expected future dividends are*

$$D_t = D_0(1+g)^t, t = 1, 2, ... \tag{2.3}$$

Then the fundamental value is

$$P_0^* = \frac{D_1}{r-g} \tag{2.4}$$

The formula is called the constant-growth DDM or the Gordon growth model (GGM). We can also write GGM in the following formula.

$$P_t^* = \frac{D_t(1+g)}{r-g} \tag{2.5}$$

In a simplified market, the Gordon Growth model can be applied to calculate the fundamental value for an individual asset. If the expected return and growth rate of dividends are constant, we call it the static Gordon Growth Model (GGM). (Heaton and Lucas, 1999) used static Gordon Growth model to determine the rational valuation of stock prices. (Fama and French, 2002) used static Gordon Growth model (GGM) to estimate the risk premium and provide empirical evidence on the decrease of the risk premium. If the expected return and growth rate of dividends are time-varying, we call it the dynamic Gordon Growth Model (GGM). There exist many variables in dynamic GGM to estimate and thus applying dynamic GGM can result in over-fitting.

2.2 Mean reversion

The mechanism of mean-reversion is a powerful force. For stocks, the stock price may swing away from its fundamental valuation, but revert back in the long run. (Poterba and Summers, 1988; Kim et al., 1991; Miller et al., 1994) and Fama and French (1988) have documented the mean reversion pattern of stock prices. (Manzan, 2007), Bail, Demirats and Levy (2008) had documented on research in mean reversion for stock prices.

2.3 Bayesian statistical methods

2.3.1 Product rule, sum rule and Bayes rule

$p(.|.)$ denotes a conditional probability distribution and $p(.)$ denotes a marginal distribution. The same notation is used for continuous density functions and discrete probability mass functions. Here the terms 'distribution' and 'density' can interchange.

We give the three rules which are the basic properties in Bayesian statistics (Bishop, 2006). Given two random variables x, y and the joint probability distribution $p(x, y)$ on these two variables, the product rule is

$$p(x, y) = p(x)p(y|x) = p(y)p(x|y), \tag{2.6}$$

the sum rule is

$$p(x) = \int p(x, y)dy, \quad y \text{ is continuous}, \tag{2.7}$$

$$p(x) = \sum_{y} p(x, y), \quad y \text{ is discrete}, \tag{2.8}$$

and the Bayes rule is

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)}. \tag{2.9}$$

Note, Bayes rule describes the relationship between the conditional probability $p(y|x)$ and the conditional probability $p(x|y)$.

Given a set of random variables $x_1, x_2, ..., x_n$, the product rule can be applied consectively and we can yield

$$p(x_1, x_2, ..., x_n) = p(x_1) \prod_{i=2}^{n} p(x_i|x_{i-1}, ..., x_1). \tag{2.10}$$

By the product rule, the joint probability distribution for parameters θ and observed data X can be written as a product of two distributions that are the prior distribution $p(\theta)$, which is not conditioned on the observations and is subjective uncertainty from the beliefs, and the likelihood function $p(X|\theta)$, which is evaluated for the observed data X and can be viewed as a function of the parameters θ respectively. That is,

$$p(X, \theta) = p(\theta)p(X|\theta). \tag{2.11}$$

By Bayes rule, we can yield the posterior distribution, which is commonly defined as the probability distribution function (pdf) of parameters θ conditioned on the observed data X,

$$p(\theta|X) = \frac{p(X, \theta)}{p(X)} = \frac{p(\theta)p(X|\theta)}{p(X)}. \tag{2.12}$$

Since the factor $p(X)$ which does not depend on parameters θ, with fixed X, can be considered a constant, then we can omit it and yield the unnormalized posterior distribution.

$$p(\theta|X) \propto p(\theta)p(X|\theta). \tag{2.13}$$

2.3.2 Bayesian framework

(Ellison, 2004) presents that Bayesians and frequentists differ in their definition of probability and in their treatment of model parameters as random variables or estimates of true value.

Bayesian data analysis is to make inferences from data using probability models for all related variables which are observed, unobserved, parameters, etc. The essential characteristic of Bayesian methods is the explicit use of probability of quantifying uncertainty. The process of Bayesian data analysis can be simplified to three steps. The first step is to construct a full probability model which should be consistent with practical problems. Normally, we need to give some assumptions to simplify the practical problems in order to construct a reasonable, but not very complicated model.

The second step is to calculate the posterior distribution which is the conditional probability distribution of the unobserved data, given the observed data. In this step, we always meet two practical difficulties in implementation. One is how to choose a prior which is largely application-specific. There are four kinds of priors which are uninformative priors, Jeffreys priors, informative priors and conjugate priors. It is common to use uninformative priors as part of any analysis to provide a kind of baseline. Another question is how to simplify the calculation of the posterior distribution. It is difficult to compute the posterior distribution, except for few special cases, e.g., Gaussian likelihood priors. We can simplify the calculation by a tool called conditional independence (CI), see Subsection 2.3.3 .

The final step is to evaluate whether the model fits the data or not, is reasonable or not etc. If necessary, we can improve the model and repeat the three steps.

2.3.3 Conditional independence, Graph and Bayesian Network

Conditional independence (CI) (Dawid,1980) is an important concept and plays an important role for probabilistic models by simplifying the structure of a model and the computations to perform inference.

Definition 2.2 (Conditional Independence, CI)**.** Given three variables x, y, z, if learning the values of the variable y does not provide additional information about the variable x, given the variable z, then the conditional distribution of x, given y and z is

$$p(x|y, z) = p(x|z). \tag{2.14}$$

The variables x and y are called conditionally independent given z, which is written

$$x \perp y|z. \tag{2.15}$$

Note that Eq.(2.14) must hold for all possible values of the variables, not just for some values. By the definition of conditional independence, it is difficult to find potential conditional independence properties. However, we can read directly from a graph to find the existing conditional independence. We introduce the graph (Bishop, 2006) as follows.

Definition 2.3. A graph comprises nodes connected by edges. A directed graph comprises nodes connected by directed links. If we walk through the directed graph by following the arrows and we never walk in a circle, then the graph is called a directed acyclic graph (DAG). For a directed graph, if a directed link is from node x to node y, then node x is called a parent of node y and node y is called a child of node x, and if there is a directed path from node x to node y on which each step follows the arrows, then node y is called a descendant of node x.

We denote the parents of node x, the children of node x, and the descendant of node x by $pa(x)$, $ch(x)$, and $desc(x)$, respectively.

Note that a graphical model (GM) is a way to represent a joint distribution. In a graphical model, each node represents a random variable and the edges represent probabilistic relationships between variables. The terms node and variable can interchange.

For a directed graph, there exist three structures which we are interested in: chain structures, fork structures and collider structures. These three structures are used to discuss conditional independences. To easily illustrate these three structures, we only consider three variables x, y and z.

A chain structure is shown by$x \rightarrow z \rightarrow y$, which satisfies $x \perp y|z$, and node z is called head-to-tail, a fork structure is shown by $x \leftarrow z \rightarrow y$, which satisfies $x \perp y|z$, and node z is called tail-to-tail, and a collider structure is shown by $x \rightarrow z \leftarrow y$,which satisfies $x \perp y|\emptyset$, and node z is called head-to-head.

In a graph with multiple nodes, we use the concept of d-separation (Lauritzen, 1996) to find potential conditional independences. In the thesis, we only focus on the directed acyclic graph (DAG).

Definition 2.4 (D-separation). In a directed acyclic graph, X, Y and Z are arbitrary nonintersecting sets of nodes. For all possible paths from any node in X to any node in Y, the path is said to be blocked if the path includes a node such that either

(a) the node is either tail-to-tail or head-to-tail , and the node is in the set Z, or

(b) the node is head-to-head, and both the node and all of its descendants are not in the set Z.

If all paths are blocked, then X is said to be d-separated from Y by Z, and the distribution over all variables in the graph satisfy $X \perp Y|Z$.

Definition 2.5 (Directed graphical model, DGM)**.** A directed graphical model (DGM) is a GM whose graph is a DAG. In DGM, each node represents a random variable, the arrows represent a set of conditional probability distributions, and for any nonintersecting sets X, Y, Z, $X \perp Y|Z \Leftrightarrow X$ is d-separated from Y given Z. DGM is also known as Bayesian Network (BN).

Bayesian Network is the integration of probability with graph theory. Bayesian Network is efficient in representing and evaluating complex probabilistic dependence structures. (Khakzad et al., 2011) present that a Bayesian Network has the ability to update probabilities and is helpful to incorporate variables and expert opinion in analysis.

Theorem 2.3 (Local Markov condition)**.** *In a DGM, each variable is independent of all its non-descendants except for its parents, conditional on its parents. That is, for an arbitrary variable x in DGM,*

$$x \perp (non - desc(x) \setminus pa(x)|pa(x)) \tag{2.16}$$

Proof. See (Cowell et al., 1999).

Combing the product rule and Theorem 2.3, Corollary 2.4 follows.

Corollary 2.4. *In a DGM, there exist n variables $x_1, x_2, ..., x_n$. The joint probability distribution over all variables can be expressed by*

$$p(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p(x_i|pa(x_i)). \tag{2.17}$$

To understand the relationship between a directed graph model and the joint probability distribution, we give an example. We consider a joint probability distribution $p(x, y, z)$ over three variables x, y and z which is

$$p(x, y, z) = p(z|x, y)p(y)p(x). \tag{2.18}$$

Next, we express the joint probability distribution in a graphical model. First, we introduce three nodes which represent three variables x, y and z. Second, we draw the arrows for each conditional distribution. For the term $p(z|x, y)$, we draw two directed links from nodes x and y to node z, respectively. For the two terms $p(x), p(y)$, there are no incoming links. The result is the graph shown in Figure 2.1.

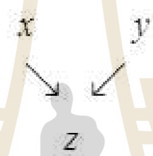

Figure 2.1 An example to express the joint probability distribution in a graphical model

Finally, by Corollary 2.4 and Figure 2.1, we write the joint probability distribution

$$\begin{aligned} p(x, y, z) &= p(z|pa(z))p(y|pa(y))p(x|pa(x)) \\ &= p(z|x, y)p(y)p(z), \end{aligned} \tag{2.19}$$

which is the equivalent to Eq.(2.18).

2.4 Parameter estimation

In Bayesian Network, as in other standard statistical models, there is a set of parameters which are conditional probabilities whose values may be known or unknown. To calculate the posterior distribution, we need to estimate unknown parameters first.

To estimate the parameters, there are three common statistical approaches: Maximum likelihood (ML), Maximum a Posteriori (MAP) and Exact Bayesian Inference. Exact Bayesian Inference is to calculate the posterior distribution exactly and is using the observations to update a prior distribution. It provides a good way to implement Ockhams razor, but it is normally intractable. Maximum a posteriori estimation is to maximize the posterior distribution which is related to likelihood and priors to estimate parameters. It is often more tractable than Exact Bayesian Inference. Maximum likelihood estimation is to maximize the likelihood function to estimate parameters. It is equivalent to MAP estimation with a uniform prior.

The expectation maximization (EM) algorithm is a general technique to find maximum likelihood solutions for probabilistic models with latent variables (Dempster et al., 1977). Maximum a Posteriori (MAP) solutions can also be solved by expectation maximization (EM) algorithm. In the thesis, we focus on the model with latent variables.

Definition 2.6. For a probabilistic model, we denote the set of all observed data, the set of all latent variables and the set of parameters by X, Z and θ, respectively. We call X the incomplete data set, the likelihood function for which is $p(X|\theta)$, whereas we call X, Z the complete data set, the likelihood function for which is the joint distribution $p(X, Z|\theta)$.

In practice, we are normally given the incomplete data X, not the complete data set X, Z and then we cannot use the complete data likelihood. Thus, our goal is to maximize the likelihood function $p(X|\theta)$ and we shall use the maximization of this log likelihood $\ln p(X|\theta)$.

Proposition 2.5. $p(X|\theta) = \dfrac{p(X, Z|\theta)}{p(Z|X, \theta)}$.

We introduce a distribution $q(Z)$ which is defined over the latent variables and satisfies $0 < q(Z) \leq 1, \sum_Z q(Z) = 1$. The knowledge for the latent variables Z

is given only by the posterior distribution $p(Z|X,\theta)$. We consider the log likelihood expected value under the posterior distribution of the latent variables. That is,

$$\ln p(X|\theta) = \sum_{Z} q(Z) \ln p(X|\theta). \tag{2.20}$$

By Proposition 2.5,

$$\ln p(X|\theta)) = \sum_{Z} q(Z) \ln \frac{p(X,Z|\theta)}{q(Z)} - \sum_{Z} q(Z) \ln \frac{p(Z|X,\theta)}{q(Z)}. \tag{2.21}$$

Definition 2.7. We set

$$\iota(q,\theta) = \textstyle\sum_Z q(Z) \ln \dfrac{p(X,Z|\theta)}{q(Z)},$$

$$KL(q||p) = -\textstyle\sum_Z q(Z) \ln \dfrac{p(Z|X,\theta)}{q(Z)},$$

where $q = q(Z), p = p(X,Z|\theta)$.

By Definition 2.7, the following decomposition holds.

$$\ln p(X|\theta)) = \iota(q,\theta) + KL(q||p). \tag{2.22}$$

Theorem 2.6. *The Kullback-Leibler divergence satisfies* $KL(q||p) \geq 0$*, with equality if and only if* $q = p$ *.*

Proposition 2.7. $\ln p(X|\theta)) \geq \iota(q,\theta)$*.*

Definition 2.8 (E-step)**.** We set

$$q^{(j)} = p^{(j-1)} \triangleq p(Z|X,\theta^{(j-1)}), j = 2,3,...,J \tag{2.23}$$

This is called E-step.

Definition 2.9 (M-step)**.** We set

$$\theta^{(j)} = \arg\max_{\theta\in\Theta}(\iota(\theta, q^{(j)})), j = 2,3,...,J \tag{2.24}$$

where Θ is a pre-defined parameter space. This is called M-step.

Theorem 2.8. *There exist two sequences* $\left\{q^{(j)}\right\}_{j=1}^{\infty}, \left\{\theta^{(j)}\right\}_{j=1}^{\infty}$ *such that*

$$\ln p(X|\theta^{(j)}) \geq \ln p(X|\theta^{(j-1)}), j = 2, 3, ..., J \tag{2.25}$$

By Theorem 2.8, the sequence $\left\{\ln p(X|\theta^{(j)})\right\}_{j=1}^{\infty}$ is an increasing sequence and since $\ln p(X|\theta^{(j)}) \leq 0$, then the sequence $\left\{\ln p(X|\theta^{(j)})\right\}_{j=1}^{\infty}$ converges. We can prove that the EM algorithm which does indeed maximize the likelihood. Note that, in the M-step, we can simplify the equation as follows:

$$\max_{\theta}(\iota(\theta, q^j)) = \max_{\theta}(\sum_{Z} q^{(j)}(Z) \ln p(X, Z|\theta)) \tag{2.26}$$

Definition 2.10. We set

$$Q(\theta; \theta^{(j-1)}) = \sum_{Z} q^{(j)}(Z) \ln p(X, Z|\theta), j = 2, 3, ..., J \tag{2.27}$$

where $q^{(j)}(Z) = p(Z|X, \theta^{(j-1)})$. Then, in the M-step,

$$\theta^{(j)} = \arg\max_{\theta \in \Theta}(Q(\theta; \theta^{(j-1)}), j = 2, 3, ..., J \tag{2.28}$$

The general EM algorithm in MAP case

Input: model assumption, $X, \Theta, \theta^{(1)}, q^{(1)}$

Repeat:

$\theta^{(j)} = \arg\max_{\theta \in \Theta}(Q(\theta; \theta^{(j-1)}) + \ln(p(\theta))), j = 2, 3, ..., J$

$j \to j + 1$

Until converges.

Output: θ_{MAP}.

2.5 Normality test for a distribution

For fitting the model or analyzing the model, data testing is necessary. We often assume that the data is to fit a normal distribution. In fact, the most widely data testing is normality test which is used to determine if a data set is well-modeled by a normal distribution and to compute how likely it is for a random variable underlying the data set to be normally distributed.

Normality tests are essentially a form of model selection and analysis, and can be interpreted by descriptive statistics, frequent statistics. These test require a relatively large sample size. If the sample size is not enough, bootstrap is useful, developed in (Efron, 1979). The key idea is that the observed data set is a random sample drawn from the actual probability distribution. The random variables are drawn from their observed distribution, which is the best estimate of the actual distribution, see (Walter, 2004). Additionally, the bootstrap is used in Subsection 4.3.

In descriptive statistics, one informal approach to testing normality is to compare a histogram of the sample data to a normal probability curve. The empirical distribution of the data (the histogram) should be bell-shaped and resemble the normal distribution.

In frequentist statistics, we have two hypotheses as follows for the normal distribution. We use statistical hypothesis testing to determine data are tested against the null hypothesis..

The two hypotheses for the frequentist statistics test for the normal distribution are given below:

- H0: The data follows the normal distribution,
- H1: The data does not follow the normal distribution.

There are several approaches to normality test, which are Anderson Darling Test, Kolmogorov-Smirnov Test, etc.

The Anderson-Darling statistic is given by the following formula:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(X_i) + \ln(1 - F(X_{n-i+1}))], \qquad (2.29)$$

where n is sample size, $F(X)$ is the cumulative distribution function for the specified distribution and i is the i^{th} sample when the data is sorted in rising order.

The K-S statistic quantifies the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. This test can be modified to serve as a goodness of fit test. In this case, samples are compared with a standard normal distribution by setting the mean and variance of the reference distribution equal to the sample estimates.

CHAPTER III

MODELING STOCK PRICING DYNAMICS BASED ON PRICE EARNING RATIO

On a daily investment decision in a security market, the price earnings (PE) ratio is one of the most widely applied methods being used as a firm valuation tool by investment experts. In this chapter, we focus on applying a Bayesian statical analysis to formalize the process of stock valuation using the PE ratio. We apply the powerful framework of dynamic Bayesian network to model the valuation process.

3.1 Background of fundamental investment based on Price Earning (PE) Ratio

The core idea of the PE ratio valuation method is simply that the value of the firm (and hence the value of its stock) is directly proportional to the annual earnings of the company, i.e. for each firm i,

$$P_i^* = PE_i^* \times E_i \tag{3.1}$$

where P_i^* denotes the value of firm i , E_i denotes the firm's current annual earnings and PE_i^* is the firm's appropriate PE ratio, usually assumed to be a constant (at least for some period of time). Here, the annual earning is defined by the summation of the latest four quarterly earnings. The earnings information of each firm listed in the stock market is normally available to all investors, i.e., it is observable.

The PE ratio, intuitively, can be thought of as a premium of an individual firm, i.e., given the same earning for two firms, the firm with higher PE ratio is considered to be of higher value. Conceptually, the appropriate PE ratio of each firm is usually determined by experts using business and financial accounting factors such as debt burden, cash flow, growth rate, business risk, etc. There exists an alternative approach to estimate the PE ratio called the relative approach (Damodaran, 2012) which still requires experts to select a group of similar firms altogether, and the ratio is heuristically calculated from this group. To summarize, the current best practice for PE ratio estimation is to be heuristically calculated by experts or experienced investors.

Once we get the PE ratio, we can simply calculate the firm value, often called intrinsic value, by Eq.(3.1). A simple trading strategy is to compare the firm value with a market price of the firm.

Strategy A: if the firm's value is higher than its market price by some threshold, it is considered to be at low price, so that we can buy the firm's stock. We expect to sell it later when its market price is higher than the firm's intrinsic value by some threshold.

It is important to note that the philosophy of this trading strategy is that the market price is not always equal to the value of the firm. We can observe that the price of the firm's stock changes almost every working day in a stock market. In contrast, by Eq.(3.1), the firm's value will not change in a short time period provided that there is no new announcement on annual earnings in that period. There has been a long controversy about this price vs. value issue (Mark, 2011), but it is beyond the scope of this thesis. In any cases, it is a fact that there exists a large group of individual investors namely fundamentalists employing the PE ratio as their main tool. Instead of solely relying on expert opinions, the goal of this part is to support that group of investors to systematically determine the appropriate PE ratio, by the method of Bayesian statistical analysis which is able to formally combine information from historical data with expert beliefs.

Finally, we emphasize that there is another quantity called an observed PE calculated from firm's current market price divided by its earnings (note again difference between value P_i^* and price P_i), that is,

$$observed\ PE_i = P_i/E_i, \tag{3.2}$$

where P_i is the current market price of a firm i. In Bayesian analysis of the PE ratio, it is important to distinguish between the observed PE_i (changing everyday due to changes of P_i) and PE_i^*. Here we will call PE_i^* as the fundamental PE ratio. The reason behind this name is the following: only the group of fundamentalists believe that the quantity P_i^* , or firm value, can be calculated by Eq.(3.1). Therefore, they usually call P_i^* as the fundamental price or fundamental value, and so PE_i^* as fundamental PE. To them, there exist various kinds of investors in the market: some are rational and some are irrational. The current market price P_i and hence, too, the observed PE_i can fluctuate from the fundamental price P_i^* by actions of those irrational investors. We shall formally model this argument in Section 3.2.

3.2 Statistical model of stock price dynamics

3.2.1 Motivation of statistical modelling: behavioral volatility

In Section 3.1, we mentioned about the fundamentalists' belief that market price of a security may not equal to its fundamental value. Why does a stock price deviate from its fundamental price? Works on behavioral finance (Szyszka, 2013) found much evidence to this question. For example, researchers argue that there are noise traders in the market who tend to make irrational actions so that the price moves away from its value (Black, 1986; De Long et al., 1990; Hommes, 2013). (Bondt and Thaler, 1985) found that some investors cannot process new information correctly and so *overreact* to new information. What is worse, information which investors overreact to is unconfirmed (Bloomfield et al., 2000) or unreliable (Pound and Zeckhauser, 1990; Tumarkin and Whitelaw, 2001) or even unimportant (Rashes, 2001; Cooper et al., 2001). Also, investors who consult experts may not get much helpful advice since security analysts tend to be overoptimistic (Dechow, Hutton, and Sloan, 2000) and have conflict of interest (Cowen, Groysberg, and Healy, 2006). Finally, it is well known that even rational investors in the market cannot immediately eliminate this irrational pricing due to limit of arbitrage (Shleifer and Vishny, 1997). All the effects mentioned here are able to temporarily move a stock price away from its value for a period of time. This is what we call behavioral volatility. The effects continue until either they are cancelled out, or rational investors finally eliminate this mispricing. This reversion phenomena is called mean reversion in the literature.

3.2.2 Dynamic Bayesian Network (DBN) of stock price movement

Our model simplifies and formalizes the observations described in Subsection 3.2.1. We divide the temporary effects which cause mispricing into two categories: (1) short-term effects: mispricing effects which last about a few days, e.g. effects caused by noise trading or overreaction to unreliable information and (2) medium-term effects : mispricing effects which last several weeks or months, e.g. effects caused by reaction to unconfirmed information which may take time to confirm, or overoptimistic prediction of analysts which may take time to prove. Mathematically, the relation between market price and its fundamental value can be described as the following equation. To simplify the equation, since we consider only one firm at a time, we now replace the firm-index subscript i with a time-index subscript t to emphasize the dynamic relationship between price and its fundamental value.

$$P_t = P_t^*(1 + z_t)(1 + \varepsilon_t) \tag{3.3}$$

where

(a) z_t is a random variable modeling the medium-term noisy effects. To make its effects persist for a period of time, we model z_t as a Markov chain.

(b) ε_t is a random variable for the short-term noisy effects which is modeled by a Gaussian random noise, $\varepsilon_t \sim N(0, \sigma^2)$.

Assuming PE^* as a constant for the period which we observed, and following Eq.(3.1) of Section 3.1, we have

$$P_t = PE^* E_t(1 + z_t)(1 + \varepsilon_t) \tag{3.4}$$

and, therefore, we get the relationship between the fundamental PE and the observed PE:

$$P_t/E_t = PE^*(1 + z_t)(1 + \varepsilon_t) \tag{3.5}$$

Note that our model is suitable only for a firm with positive earnings $E_t > 0$. Fortunately, most firms satisfy this criterion. Eq.(3.5) is central to our idea and can be visualized as shown in Figure 3.1. In Figure 3.1, the plot is the observed PE in time series. The red dashed line shows PE^* of the firm, while the green line illustrates the effect of medium-term effect z_t to PE^*. The blue line illustrates the observed PE which is affected by both the medium-term and short-term mispricing effects.

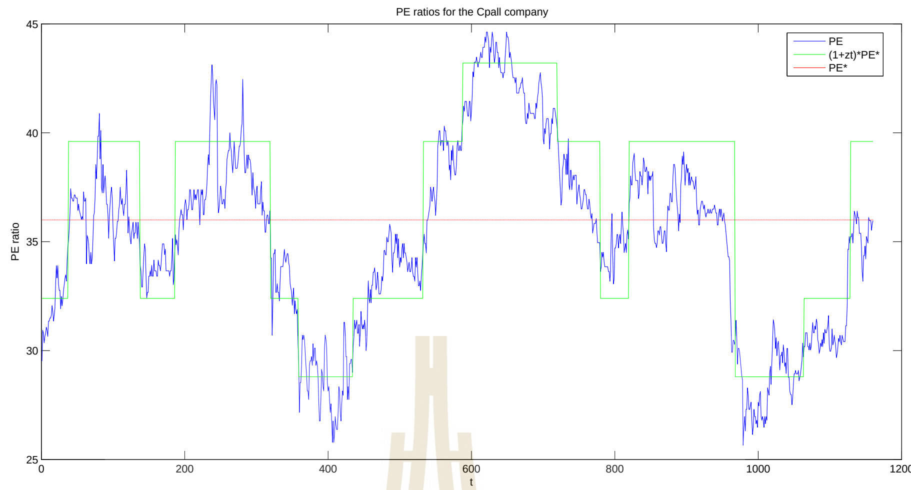

Figure 3.1 Illustration of our main idea described by Eq. (3.5).

We can mathematically simplify Eq.(3.5) further

$$\ln(P_t/E_t) = \ln(PE^*(1+z_t)) + \ln(1+\varepsilon_t) \tag{3.6}$$

Since ε_t is usually small, it can be approximated by $\ln(1+\varepsilon_t) \approx \varepsilon_t$, and denote $y_t = \ln(P_t/E_t)$, we then have

$$y_t = \ln(PE^*(1+z_t)) + \varepsilon_t \tag{3.7}$$

Note that as explained in Section 3.1, y_t is an observable quantity, whereas PE^* and z_t are unobservable, i.e. they are hidden state or latent variables. Note that these are two different types of latent variables, i.e. PE^* is constant and z_t is time-varying. Thus, Eq.(3.7) is different from standard state-space and graphical models such as Hidden Markov Models or Linear State Space Model (Bishop, 2006). The graphical model of our proposed stock price dynamic has three layers as represented in Figure 3.2. In our case, where the model is temporal, the graphical model framework is also called dynamic Bayesian network (DBN). The main advantage of DBN is its ability to encode conditional independent properties, and hence to simplify probabilistic inference (K. P. Murphy, 2012). Another advantage of this framework is that expert knowledge can be integrated in the model naturally as shown in the next section.

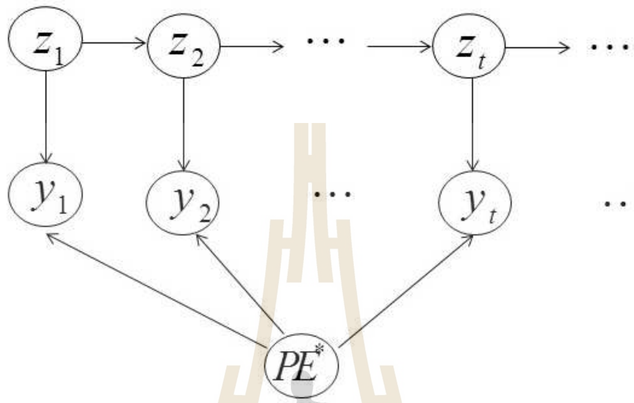

Figure 3.2 The proposed model represented by dynamic Bayesian network (DBN). y_t is an observable quantity, while PE^* and $\{z_t\}$ are unobservable.

To derive mathematical equations for inference and parameter estimation in the DBN framework, we shall assume that all latent random variables are discrete: $z_t \in \{a_1, ..., a_M\}$, $PE^* \in \{b_1, ..., b_N\}$. There are 4 main conditional independence properties of our model.

(CI1) For each $t \in 1, 2, ..., z_t \perp PE^* | \emptyset$.

(CI2) For all $i, j, k \in 1, 2, ...,$ if there exist $k, i < k < j$, then $z_i \perp z_j | z_k$.

(CI3) For each $i, j, k \in 1, 2, ...,$ if there exist $k, i \leq k < j$ or $j < k \leq j$ then $y_i \perp z_j | z_k, PE^*$.

(CI4) For all $i, j, k \in 1, 2, ...,$ if there exist $k, i < j, i \leq k \leq j$, then $y_i \perp y_j | z_k, PE^*$.

Furthermore, we have to set up the conditional probability distribution function for each node given its parents. We define the conditional probability distribution functions of all nodes as follows:

The transition probability distribution function:

Let $i, m \in \{1, ..., M\}$, $t \in \{2, 3, ...\}$

$$p(z_t = a_i \mid z_{t-1} = a_m) \triangleq w_{im}. \tag{3.8}$$

Note that $0 \leq w_{im} \leq 1$, $\sum_{m=1}^{M} w_{im} = 1$. The matrix $\mathbf{W} = (w_{im})_{M \times M}$ is called a transition matrix, i.e. $\{z_t\}$ is a Markov chain.

The emission probability distribution function (pdf):

For all $m \in \{1, ..., M\}$, $n \in \{1, ..., N\}$,$t \in \{1, 2, ...\}$

$$p(y_t | z_t = a_m, PE^* = b_n) \triangleq \phi_{mn}(y_t). \tag{3.9}$$

By Eq.(3.6), $\phi_{mn}(y_t) = N(\ln(b_n(1 + a_m)), \sigma^2)$. The matrix $\Phi_t = (\phi_{mn})_{M \times N}$ is called an emission matrix at period t.

The inital probability distribution functions:

For each $m \in \{1, ..., M\}$,

$$u_m \triangleq p(z_1 = a_m) \tag{3.10}$$

where $0 \leq u_m \leq 1$ and $\sum_{m=1}^{M} u_m = 1$.

For each $n \in \{1, ..., N\}$,

$$v_n \triangleq p(PE^* = b_n) \tag{3.11}$$

where $0 \leq v_n \leq 1$ and $\sum_{n=1}^{N} v_n = 1$. The vectors $\mathbf{u} = (u_m)_M$ and $\mathbf{v} = (v_n)_N$ are called initial vectors.

Therefore, in this Bayesian framework, the set of model parameters is $\theta = \{\mathbf{W},\mathbf{u},\mathbf{v}, \sigma^2\}$ and our parameters space is

$$\begin{aligned}\Theta = \{\theta | 0 \leq u_m \leq 1, \sum_{m=1}^{M} u_m = 1, 0 \leq v_n \leq 1, \sum_{n=1}^{N} v_n = 1,\\ 0 \leq w_{im} \leq 1, \sum_{m=1}^{M} w_{im} = 1, \sigma > 0\}.\end{aligned} \tag{3.12}$$

If we know all parameters, we can make an inference by deriving inference equations based on the forward-backward algorithm. If the parameters are unknown, we have to estimate them first. In the next section, we derive the estimation procedures based on Maximum a Posteriori (MAP) and Expectation-Maximization (EM) algorithms and we show how to derive both the inference and parameter estimation algorithms.

3.3 Bayesian inference on the DBN of stock price dynamic

As explained in the previous sections, our goal is to make an inference on PE* so that we can estimate the fundamental price of a stock. In Chapter IV, we will show that estimations of $\{z_t\}$ are also useful in investment. To infer the values of these two latent variables, similar to Hidden Markov Models (HMM) and Linear State Space Model (LSSM)(Bishop, 2006; K. P. Murphy, 2012), we need to derive equations in two steps. First, the inference algorithms with known parameters; second, the parameter estimation algorithms given that parameters are unknown. However, because there are two types of latent states as explained in the previous section, our graphical model shown in Figure 3.2 is more sophisticated than HMM and LSSM. In this section, we show the new equations for both inference tasks. To simplify the notation, we use notation x_1^T to denote$\{x_1, ..., x_T\}$.

3.3.1 Inference with known parameters

Suppose θ is known, together with the observed data y_1^T. Similar to HMM, in order to estimate the latent states of z_1^T and PE^*, we need to find recurrent formulas to calculate two quantities: the *filtering* probabilities $p(z_T, PE^*|y_1^T, \theta)$ and the *smoothing* probabilities $p(z_t, PE^*|y_1^T, \theta)$, $t \in \{1, ..., T-1\}$. To keep our formulas simple, we will omit writing θ in the probability notations, e.g., we simply write $p(z_T, PE^*|y_1^T)$ for filtering.

3.3.1.1 Filtering probability distribution function

The filtering formula is to estimate conditional joint probabilities of the *most recent* medium-term effect $z_T = a_m$ and $PE^* = b_n$ given all the observed variables. For all $t \in \{1, ..., T\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$\alpha_{tmn} \triangleq p(z_t = a_m, PE^* = b_n|y_1^t). \tag{3.13}$$

First, for all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$, we calculate α_{1mn},

$$\begin{aligned}
\alpha_{1mn} &\triangleq p(z_t = a_m, PE^* = b_n|y_1) \\
&= \frac{p(y_1|z_1 = a_m, PE^* = bn)p(z_1 = a_m, PE^* = b_n)}{p(y_1)} \\
&\propto p(y_1|z_1 = a_m, PE^* = bn)p(z_1 = a_m, PE^* = b_n) \\
&= p(y_1|z_1 = a_m, PE^* = bn)p(z_1 = a_m)p(PE^* = b_n) \\
&= \phi_{mn}(y_1)u_m v_n.
\end{aligned} \tag{3.14}$$

Then, the initial equation of the recurrent formula can be derived: $\alpha_{1mn} \propto \phi_{mn}(y_1)u_m v_n$. For all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$\alpha'_{1mn} = \phi_{mn}(y_1)u_m v_n. \tag{3.15}$$

By Eq.(3.13), for all $t \in \{1, ..., T\}$,

$$\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{tmn} = 1. \tag{3.16}$$

By Eq.(3.16), for all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\alpha_{1mn} = \frac{\alpha'_{1mn}}{\sum_{m'=1}^{M}\sum_{n'=1}^{N}\alpha'_{1m'n'}}. \tag{3.17}$$

Therefore, for all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\alpha_{1mn} = \frac{\phi_{mn}(y_1)u_m v_n}{\sum_{m'=1}^{M}\sum_{n'=1}^{N}\phi_{m'n'}(y_1)u_{m'}v_{n'}}. \tag{3.18}$$

Second, for all $t \in \{2, ..., T\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, we give the following recurrent formula.

$$\begin{aligned}
&p(z_t = a_m, PE^* = b_n|y_1^t)\\
&= p(z_t = a_m, PE^* = b_n|y_1^{t-1}, y_t)\\
&\propto p(y_t|y_1^{t-1}, z_t = a_m, PE^* = b_n)p(z_t = a_m, PE^* = b_n|y_1^{t-1})\\
&= \phi_{mn}(y_t)\textstyle\sum_{i=1}^{M} p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^{t-1})\\
&= \phi_{mn}(y_t)\textstyle\sum_{i=1}^{M} p(z_{t-1} = a_i, PE^* = b_n|y_1^{t-1})p(z_t = a_m|z_{t-1} = a_i)\\
&= \phi_{mn}(y_t)\textstyle\sum_{i=1}^{M} p(z_{t-1} = a_i, PE^* = b_n|y_1^{t-1})w_{mi}
\end{aligned} \tag{3.19}$$

Then for all $t \in \{2, ..., T\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\alpha_{tmn} \propto \phi_{mn}(y_t)\sum_{i=1}^{M} p(z_{t-1} = a_i, PE^* = b_n|y_1^{t-1})w_{mi}. \tag{3.20}$$

For all $t \in \{2, ..., T\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$\alpha'_{tmn} = \phi_{mn}(y_t)\sum_{i=1}^{M} p(z_{t-1} = a_i, PE^* = b_n|y_1^{t-1})w_{mi}. \tag{3.21}$$

Then, for all $t \in \{2, ..., T\}$, we have

$$\alpha_{tmn} \propto \alpha'_{tmn}. \tag{3.22}$$

By Eq.(3.16), for all $t \in \{2, ..., T\}$,

$$\alpha_{tmn} = \alpha'_{tmn} / \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha'_{tmn}. \tag{3.23}$$

According to Eq.(3.19) and Eq.(3.23), we can continue to calculate $(\alpha_{2mn}), ..., (\alpha_{Tmn})$ as follows and finally we can get the filtering formula α_{Tmn}.

$$\begin{pmatrix} \alpha_{111} & \cdots & \alpha_{11N} \\ \cdots & \cdots & \cdots \\ \alpha_{1M1} & \cdots & \alpha_{1MN} \end{pmatrix} \rightarrow \begin{pmatrix} \alpha_{211} & \cdots & \alpha_{21N} \\ \cdots & \cdots & \cdots \\ \alpha_{2M1} & \cdots & \alpha_{2MN} \end{pmatrix} \rightarrow \cdots \rightarrow \begin{pmatrix} \alpha_{T11} & \cdots & \alpha_{T1N} \\ \cdots & \cdots & \cdots \\ \alpha_{TM1} & \cdots & \alpha_{TMN} \end{pmatrix} \tag{3.24}$$

Note that, for all $t \in \{1, ..., T\}$, we denote

$$c_t = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha'_{tmn}. \tag{3.25}$$

In the above derivation, Bayes's rule, *conditional independent* properties (K. P. Murphy, 2012) with respect to DBN shown in Figure 3.2 and sum rule are applied consecutively to get the above result, similar to the filtering equation of HMM.

3.3.1.2 Smoothing probability distribution function

Next, the smoothing formula is to estimate conditional joint probabilities of the *any-date* $t < T$ medium-term noisy effect $z_t = a_m$ and $PE^* = b_n$ given all the observed variables.

For all $t \in \{1, ..., T-1\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$\gamma_{tmn} \triangleq p(z_t = a_m, PE^* = b_n | y_1^T). \tag{3.26}$$

For all $t \in \{1, ..., T-1\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\begin{aligned}
&p(z_t = a_m, PE^* = b_n | y_1^t, y_{t+1}^T) \\
&= \frac{p(y_{t+1}^T | y_1^t, z_t = a_m, PE^* = b_n)}{p(y_{t+1}^T | y_1^t)} p(z_t = a_m, PE^* = b_n | y_1^t) \\
&= \frac{p(y_{t+1}^T | z_t = a_m, PE^* = b_n)}{p(y_{t+1}^T | y_1^t)} \alpha_{tmn}.
\end{aligned} \tag{3.27}$$

For each $t \in \{1, ..., T-1\}$, by chain rule,

$$p(y_{t+1}^T | y_1^t) = \prod_{t'=t+1}^{T} c_{t'}. \tag{3.28}$$

For all $t \in \{1, ..., T-1\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$\beta_{tmn} \triangleq \frac{p(y_{t+1}^T | z_t = a_m, PE^* = b_n)}{p(y_{t+1}^T | y_1^t)}, \tag{3.29}$$

and

$$\beta'_{tmn} \triangleq p(y_{t+1}^T | z_t = a_m, PE^* = b_n). \tag{3.30}$$

Then, for all $t \in \{1, ..., T-1\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\beta_{tmn} = \frac{\beta'_{tmn}}{\prod_{t'=t+1}^{T} c_{t'}}, \tag{3.31}$$

and

$$\gamma_{tmn} = \alpha_{tmn} \beta_{tmn}. \tag{3.32}$$

First, we calculate $\beta_{T-1,mn}$ for all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$.

$$\beta'_{T-1,mn} = p(y_T|z_{T-1} = a_m, PE^* = b_n)$$

$$= \sum_{i=1}^{M} p(y_T, z_T = a_i|z_{T-1} = a_m, PE^* = b_n)$$

$$= \sum_{i=1}^{M} p(y_T|z_T = a_i, z_{T-1} = a_m, PE^* = b_n)p(z_T = a_i|z_{T-1} = a_m, PE^* = b_n)$$

$$= \sum_{i=1}^{M} p(y_T|z_T = a_i, PE^* = b_n)p(z_T = a_i|z_{T-1} = a_m)$$

$$= \sum_{i=1}^{M} \phi_{in}(y_T)w_{im}. \tag{3.33}$$

Then, for all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\beta'_{T-1,mn} = \sum_{i=1}^{M} \phi_{in}(y_T)w_{im}. \tag{3.34}$$

By Eq.(3.25), Eq.(3.28), Eq.(3.31) and Eq.(3.34),

$$\beta_{T-1,mn} = \frac{\beta'_{T-1,mn}}{c_T} = \frac{\sum_{i=1}^{M} \phi_{in}(y_T)w_{im}}{\sum_{m=1}^{M}\sum_{n=1}^{N} \alpha'_{Tmn}}. \tag{3.35}$$

Second, for all $t \in \{1, ..., T-2\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, we give the following recurrent relation.

$$
\begin{aligned}
\beta'_{tmn} &= p(y_{t+1}^T|z_t = a_m, PE^* = b_n)\\
&= \textstyle\sum_{i=1}^M p(y_{t+1}^T, z_{t+1} = a_i|z_t = a_m, PE^* = b_n)\\
&= \textstyle\sum_{i=1}^M p(y_{t+1}^T|z_{t+1} = a_i, z_t = a_m, PE^* = b_n)p(z_{t+1} = a_i|z_t = a_m, PE^* = b_n)\\
&= \textstyle\sum_{i=1}^M p(y_{t+1}^T|z_{t+1} = a_i, PE^* = b_n)p(z_{t+1} = a_i|z_t = a_m)\\
&= \textstyle\sum_{i=1}^M p(y_{t+1}^T|z_{t+1} = a_i, PE^* = b_n)w_{im}\\
&= \textstyle\sum_{i=1}^M p(y_{t+1}, y_{t+2}^T|z_{t+1} = a_i, PE^* = b_n)w_{im}\\
&= \textstyle\sum_{i=1}^M p(y_{t+2}^T|y_{t+1}, z_{t+1} = a_i, PE^* = b_n)p(y_{t+1}|z_{t+1} = a_i, PE^* = b_n)w_{im}\\
&= \textstyle\sum_{i=1}^M p(y_{t+2}^T|z_{t+1} = a_i, PE^* = b_n)\phi_{in}(y_{t+1})w_{im}\\
&= \textstyle\sum_{i=1}^M \beta'_{t+1,in}\phi_{in}(y_{t+1})w_{im}.
\end{aligned}
\tag{3.36}
$$

Then for all $t \in \{1, ..., T-2\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$,

$$\beta'_{tmn} = \sum_{i=1}^{M} \beta'_{t+1,in}\phi_{in}(y_{t+1})w_{im}. \tag{3.37}$$

By Eq.(3.25), Eq.(3.28), Eq.(3.31) and Eq.(3.37),

$$(\prod_{t'=t+1}^{T} c_{t'})\beta_{tmn} = \sum_{i=1}^{M}(\prod_{t''=t+2}^{T} c_{t''})\beta_{t+1,in}\phi_{in}(y_{t+1})w_{im}, \tag{3.38}$$

$$\beta_{tmn} = \frac{1}{c_{t+1}} \sum_{i=1}^{M} \beta_{t+1,mn}\phi_{in}(y_{t+1})w_{im}. \tag{3.39}$$

According to Eq.(3.36) and Eq.(3.31), we can continue to calculate $(\beta_{T-2,mn}), ..., (\beta_{1mn})$ below. Finally, combing the Eq.(3.32), we can get the smoothing formula γ_{tmn}, where $t \in \{1, ..., T-1\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$.

$$
\begin{pmatrix} \beta_{T-1,11} & \cdots & \beta_{T-1,1N} \\ \cdots & \cdots & \cdots \\ \beta_{T-1,M1} & \cdots & \beta_{T-1,MN} \end{pmatrix} \rightarrow \begin{pmatrix} \beta_{T-2,11} & \cdots & \beta_{T-2,1N} \\ \cdots & \cdots & \cdots \\ \beta_{T-2,M1} & \cdots & \beta_{T-2,MN} \end{pmatrix} \rightarrow \cdots \rightarrow \begin{pmatrix} \beta_{111} & \cdots & \beta_{11N} \\ \cdots & \cdots & \cdots \\ \beta_{1M1} & \cdots & \beta_{1MN} \end{pmatrix} \tag{3.40}
$$

The process of calculating both filtering and smoothing is shown in Figure 3.3.

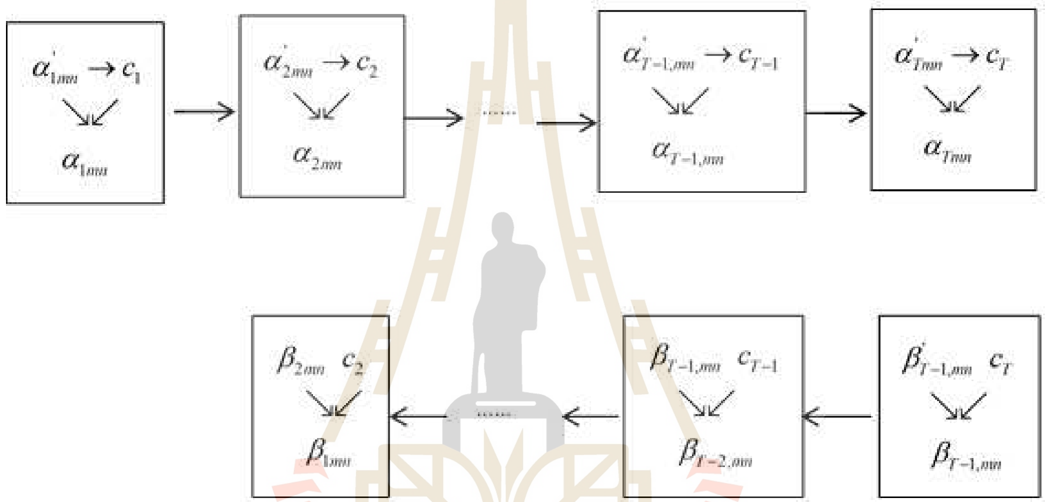

Figure 3.3 The process of calculating filtering and smoothing using forward-backward algorithm

With the derived recurrent formulas, we can get the most probable values of wanted latent variables PE^* and each z_t by using marginalization, e.g.

$$PE^* = \arg\max_{b_n} p(PE^* = b_n|y_1^T),$$

where $p(PE^* = b_n|y_1^T) = \sum_{m=1}^{M} p(z_t = a_m, PE^* = b_n|y_1^T)$. To implement both filtering and smoothing in computer program, we also need to solve the formulas for the constants appearing in the above derivations. To fulfil this task, using matrix reformulation of the above recurrent equations is the most convenient and efficient way. Below, we give the matrix formulas.

3.3.1.3 Matrix formulas of filtering and smoothing probability distribution function

To get a matrix formula for a filtering density, for all $t \in \{1, ..., T\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, we denote the matrix $\mathbf{A}_t = (\alpha_{tmn})_{M\times N}$. For the initial case, by Eq.(3.15), Eq.(3.17) and Eq.(3.25) we can have

$$\begin{pmatrix} \alpha_{111} & \cdots & \alpha_{11N} \\ \vdots & \ddots & \vdots \\ \alpha_{1M1} & \cdots & \alpha_{1MN} \end{pmatrix} = \frac{1}{c_1} \begin{pmatrix} \phi_{11}(y_1) & \cdots & \phi_{11N}(y_1) \\ \vdots & \ddots & \vdots \\ \phi_{1M1}(y_1) & \cdots & \phi_{1MN}(y_1) \end{pmatrix} \circ \left[\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_N \end{pmatrix} \right] \tag{3.41}$$

Therefore,

$$\mathbf{A}_1 = \frac{1}{c_1}\Phi_1 \circ (\mathbf{u}\mathbf{v}^T) \tag{3.42}$$

where $\mathbf{u}$ and $\mathbf{v}$ are as defined in Section 3.2. Φ_t and $\mathbf{W}$ denote the emission matrix and transition matrix, respectively, as described in Section 3.2.

For $t \in \{2, ..., T\}$, we have

$$\begin{pmatrix} \alpha_{t11} & \cdots & \alpha_{t1N} \\ \cdots & \cdots & \cdots \\ \alpha_{tM1} & \cdots & \alpha_{tMN} \end{pmatrix} = \frac{1}{c_t} \begin{pmatrix} \phi_{11}(y_t) & \cdots & \phi_{1N}(y_t) \\ \cdots & \cdots & \cdots \\ \phi_{M1}(y_t) & \cdots & \phi_{MN}(y_t) \end{pmatrix} \circ \left[\begin{pmatrix} w_{11} & \cdots & w_{1M} \\ \cdots & \cdots & \cdots \\ w_{M1} & \cdots & w_{MM} \end{pmatrix} \begin{pmatrix} \alpha_{t-1,11} & \cdots & \alpha_{t-1,1N} \\ \cdots & \cdots & \cdots \\ \alpha_{t-1,M1} & \cdots & \alpha_{t-1,MN} \end{pmatrix} \right] \tag{3.43}$$

Therefore,

$$\mathbf{A}_t = \frac{1}{c_t}\Phi_t \circ (\mathbf{W}\mathbf{A}_{t-1}), \quad t > 2 \tag{3.44}$$

where $\circ$ denotes the *entrywise* (or Hadamard) product of the matrix.

To get a matrix formula for a smoothing density, for all $m \in \{1, ..., M\}, n \in \{1, ..., N\}$, we denote the matrix $\mathbf{B}_t = (\beta_{tmn})_{M\times N}, t < T$. For the case $t = T - 1$, by Eq.(3.35), we have

$$\begin{pmatrix} \beta_{T-1,11} & \cdots & \beta_{T-1,1N} \\ \cdots & \cdots & \cdots \\ \beta_{T-1,M1} & \cdots & \beta_{T-1,MN} \end{pmatrix} = \frac{1}{c_T} \begin{pmatrix} w_{11} & \cdots & w_{1M} \\ \cdots & \cdots & \cdots \\ w_{M1} & \cdots & w_{MM} \end{pmatrix}^T \begin{pmatrix} \phi_{11}(y_T) & \cdots & \phi_{1N}(y_T) \\ \cdots & \cdots & \cdots \\ \phi_{M1}(y_T) & \cdots & \phi_{MN}(y_T) \end{pmatrix} \tag{3.45}$$

Therefore,

$$\mathbf{B}_{T-1} = \frac{1}{c_T}\mathbf{W}^T\Phi_T. \tag{3.46}$$

For $t \in \{1, ..., T-2\}$, by Eq.(3.39), we have

$$\begin{pmatrix} \beta_{t11} & \cdots & \beta_{t1N} \\ \cdots & \cdots & \cdots \\ \beta_{tM1} & \cdots & \beta_{tMN} \end{pmatrix} = \frac{1}{c_{t+1}} \begin{pmatrix} w_{11} & \cdots & w_{1M} \\ \cdots & \cdots & \cdots \\ w_{M1} & \cdots & w_{MM} \end{pmatrix}^T \left[\begin{pmatrix} \phi_{11}(y_t+1)) & \cdots & \phi_{1N}(y_t+1) \\ \cdots & \cdots & \cdots \\ \phi_{M1}(y_t+1) & \cdots & \phi_{MN}(y_t+1) \end{pmatrix} \circ \begin{pmatrix} \beta_{t+1,11} & \cdots & \beta_{t+1,1N} \\ \cdots & \cdots & \cdots \\ \beta_{t+1,M1} & \cdots & \beta_{t+1,MN} \end{pmatrix} \right]. \tag{3.47}$$

Therefore,

$$\mathbf{B}_t = \frac{1}{c_{t+1}}\mathbf{W}^T(\Phi_{t+1} \circ \mathbf{B}_{t+1}), \quad t \in \{1, ..., T-2\}. \tag{3.48}$$

For $t \in \{1, ..., T-1\}$, by Eq.(3.46)and Eq.(3.48), we have

$$\mathbf{B}_t = \begin{cases} \dfrac{1}{c_T}\mathbf{W}^T\mathbf{\Phi}_T, & t = T-1 \\ \dfrac{1}{c_{t+1}}\mathbf{W}^T(\mathbf{\Phi}_{t+1} \circ \mathbf{B}_{t+1}), & t \in \{1, \cdots, T-2\}. \end{cases} \tag{3.49}$$

3.3.2 Inference with unknown parameters

In general situations, θ is unknown, so only the observed data y_1^T is available. In this case, θ must be estimated first. *Expectation Maximization* (EM) is a general method capable of estimating the parameters θ in *Maximum Likelihood* and *Maximum a Posteriori* (MAP) problem settings for probabilistic models with latent variables (Laird et al., 1977). Here, we formulate our parameter estimation in the MAP setting so that expert's prior knowledge can be employed into the model. Formally, we would like to solve the following problem of maximizing the posterior probability distribution function of θ.

$$\theta_{MAP} = \arg\max_{\theta\in\Theta} p(\theta|y_1^T). \tag{3.50}$$

EM finds a solution of Eq.(3.50) by iteratively solving the following two steps with the arbitrary set of initial parameters $\theta^{(1)}$ and a prior $p(\theta)$. Iterating from $j = 1, 2, ...,$ do

E-Step: Calculate smoothing probabilities $p(z_t = a_m, PE^* = b_n|y_1^T, \theta^{(j)})$, $\forall t, m, n$

M-step: Solve the constraint maximization problem,

$$\theta^{(j+1)} = \arg\max_{\theta\in\Theta}[Q(\theta;\theta^{(j)}) + \ln p(\theta)] \tag{3.51}$$

where

$$Q(\theta;\theta^{(j)}) = E_{z_1^T, PE^*|y_1^T,\theta^{(j)}}[\ln p(y_1^T, z_1^T, PE^*|\theta)] \tag{3.52}$$

3.3.2.1 E-step calculation

EM repeats the two steps until $\theta^{(j)}$ converges. Note that EM guarantees to find a local maxima of Eq.(3.50) (Bishop, 2006). The argument in the expectation of Eq.(3.52) is simply the log-likelihood of the model:

$$\begin{aligned}\ln p(y_1^T, z_1^T, PE^*|\theta) = \sum_{t=1}^{T} \ln p(y_t|z_t, PE^*, \theta) + \sum_{t'=2}^{T} \ln p(z_{t'}|z_{t'-1}, \theta) \\ + \ln p(z_1|\theta) + \ln p(PE^*|\theta).\end{aligned} \tag{3.53}$$

According to DBN, they are simply the logarithms of the emission pdf, transition pdf and initial pdf, respectively. By equation manipulations, the expectation Eq.(3.52) can be calculated by employing the smoothing probabilities already done in the E-step. As a result, we get a closed form of Eq.(3.52). We begin to calculate the $Q(\theta;\theta^{(j)})$.

By Eq.(3.52) and Eq.(3.53),

$$
\begin{aligned}
&Q(\theta;\theta^{(j)}) \\
&= \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(y_1^T, z_1^T, PE^*|\theta) \\
&= \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)[\sum_{t=1}^{T}\ln p(y_t|z_t, PE^*, \theta) \\
&+ \sum_{t'=2}^{T}\ln p(z_{t'}|z_{t'-1},\theta) + \ln p(z_1|\theta) + \ln p(PE^*|\theta)] \\
&= \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\sum_{t=1}^{T}\ln p(y_t|z_t, PE^*, \theta) + \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\sum_{t'=2}^{T}\ln p(z_{t'}|z_{t'-1},\theta) \\
&+ \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(z_1|\theta) + \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(PE^*|\theta).
\end{aligned} \tag{3.54}
$$

where

$$q^j(z_1^T, PE^*) = p(z_1^T, PE^*|y_1^T, \theta^{(j-1)}), \tag{3.55}$$

and

$$0 \le q^j(z_1^T, PE^*) \le 1, \sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*) = 1. \tag{3.56}$$

We consider term by term in Eq.(3.54) to simplify $Q(\theta;\theta^{(j)})$. The first term is

$$\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\sum_{t=1}^{T}\ln p(y_t|z_t, PE^*, \theta). \tag{3.57}$$

$$\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*) \sum_{t=1}^{T} \ln p(y_t|z_t, PE^*, \theta)$$
$$= \sum_{t=1}^{T}\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*) \ln p(y_t|z_t, PE^*, \theta)$$
$$= \sum_{t=1}^{T}\sum_{z_1,...,z_T}\sum_{PE^*} q^j(z_1, ..., z_T, PE^*) \ln p(y_t|z_t, PE^*, \theta)$$
$$= \sum_{t=1}^{T}\sum_{z_t}\sum_{PE^*} q^j(z_t, PE^*) \ln p(y_t|z_t, PE^*, \theta)$$
$$= \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^j(z_t = a_m, PE^* = b_n) \ln p(y_t|z_t = a_m, PE^* = b_n, \theta). \tag{3.58}$$

For all $t \in \{1, ..., T\}, m \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$q^{(j)}_{tmn} \triangleq q^{(j)}_{tmn}(z_t, PE^*) \triangleq q^j(z_t = a_m, PE^* = b_n), \tag{3.59}$$

where, $q^j(z_t = a_m, PE^* = b_n) = p(z_t = a_m, PE^* = b_n|y_1^T, \theta^{(j-1)})$. Then,

$$\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*) \sum_{t=1}^{T} \ln p(y_t|z_t, PE^*, \theta) = \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^j_{tmn} \ln \phi_{mn}(y_t). \tag{3.60}$$

Since the definitions of filtering and smoothing Eq.(3.13) and Eq.(3.32) are described in Subsection 3.3.1, then by Eq.(3.13), Eq.(3.26) and Eq.(3.55),

$$q^{(j)}_{Tmn} = \alpha^{(j-1)}_{Tmn}, \tag{3.61}$$

and

$$q^{(j)}_{tmn} = \alpha^{(j-1)}_{tmn}\beta^{(j-1)}_{tmn}, t \in \{1, ..., T-1\}. \tag{3.62}$$

Thus, by Eq.(3.60), Eq.(3.61) and Eq.(3.62), the first term is

$$\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*) \sum_{t=1}^{T} \ln p(y_t|z_t, PE^*, \theta)$$
$$= \sum_{t=1}^{T-1}\sum_{m=1}^{M}\sum_{n=1}^{N} \alpha^{(j-1)}_{tmn}\beta^{(j-1)}_{tmn} \ln \phi_{mn}(y_t) + \sum_{m=1}^{M}\sum_{n=1}^{N} \alpha^{(j-1)}_{Tmn} \ln \phi_{mn}(y_T). \tag{3.63}$$

The second term is

$$\sum_{z_1^T} \sum_{PE^*} q^j(z_1^T, PE^*) \sum_{t'=2}^{T} \ln p(z_{t'} | z_{t'-1}, \theta). \tag{3.64}$$

$$\begin{aligned}
&\sum_{z_1^T} \sum_{PE^*} q^j(z_1^T, PE^*) \sum_{t'=2}^{T} \ln p(z_{t'} | z_{t'-1}, \theta) \\
=& \sum_{t'=2}^{T} \sum_{z_1^T} \sum_{PE^*} q^j(z_1^T, PE^*) \ln p(z_{t'} | z_{t'-1}, \theta) \\
=& \sum_{t'=2}^{T} \sum_{z_{t'}, z_{t'-1}} \sum_{PE^*} q^j(z_{t'}, z_{t'-1} PE^*) \ln p(z_{t'} | z_{t'-1}, \theta) \\
=& \sum_{t'=2}^{T} \sum_{m,i=1}^{M} \sum_{n=1}^{N} q^j(z_{t'} = a_m, z_{t'-1} = a_i, PE^* = b_n) \ln p(z_{t'} = a_m | z_{t'-1} = a_i, \theta).
\end{aligned} \tag{3.65}$$

For all $t' \in \{2, ..., T\}, m, i \in \{1, ..., M\}, n \in \{1, ..., N\}$, let

$$q_{t',m,i,n}^{(j)} \triangleq q_{t',m,i,n}^{(j)}(z_{t'}, z_{t'-1}, PE^*) \triangleq q^j(z_{t'} = a_m, z_{t'-1} = a_i, PE^* = b_n). \tag{3.66}$$

Then,

$$\begin{aligned}
&\sum_{z_1^T} \sum_{PE^*} q^{(j)}(z_1^T, PE^*) \sum_{t'=2}^{T} \ln p(z_{t'} | z_{t'-1}, \theta) \\
=& \sum_{t'=2}^{T} \sum_{m,i=1}^{M} \sum_{n=1}^{N} q_{t',m,i,n}^{(j)}(z_{t'}, z_{t'-1}, PE^*) \ln w_{mi},
\end{aligned} \tag{3.67}$$

where $q_{t',m,i,n}^{(j)}(z_{t'}, z_{t'-1}, PE^*) = p(z_{t'} = a_m, z_{t'-1} = a_i, PE^* = b_n | y_1^T, \theta^{(j-1)})$.

For the term $q_{t',m,i,n}^{j}(z_{t'}, z_{t'-1}, PE^*)$, we separate into two steps to calculate it.

In step 1, we calculate $q_{T,m,i,n}^{j}(z_T, z_{T-1}, PE^*)$. For all $m, i \in \{1, ..., M\}, n \in$

$\{1, ..., N\}$,

$$
\begin{aligned}
& q^{j}_{T,m,i,n}(z_T, z_{T-1}, PE^{*}) \\
=& p(z_{t'} = a_m, z_{t'-1} = a_i, PE^{*} = b_n | y_1^T, \theta^{(j-1)}) \\
=& p(z_{t'} = a_m, z_{t'-1} = a_i, PE^{*} = b_n | y_T, y_1^{T-1}, \theta^{(j-1)}) \\
=& \frac{p(y_T | z_T = a_m, PE^{*} = b_n, \theta^{(j-1)}) p(z_T = a_m, z_{T-1} = a_i, PE^{*} = b_n | y_1^{T-1}, \theta^{(j-1)})}{p(y_T | y_1^{T-1}, \theta^{(j-1)})} \\
=& \frac{\phi_{mn}^{(j-1)(y_T)}}{c_T^{(j-1)}} p(z_T = a_m, z_{T-1} = a_i, PE^{*} = b_n | y_1^{T-1}, \theta^{(j-1)}) \\
=& \frac{\phi_{mn}^{(j-1)(y_T)}}{c_T^{(j-1)}} p(z_T = a_m | z_{T-1} = a_i, PE^{*} = b_n, y_1^{T-1}, \theta^{(j-1)}) \\
& p(z_{T-1} = a_i, PE^{*} = b_n | y_1^{T-1}, \theta^{(j-1)}) \\
=& \frac{\phi_{mn}^{(j-1)(y_T)}}{c_T^{(j-1)}} p(z_T = a_m | z_{T-1} = a_i, \theta^{(j-1)}) p(z_{T-1} = a_i, PE^{*} = b_n | y_1^{T-1}, \theta^{(j-1)}) \\
=& \frac{\phi_{mn}^{(j-1)}(y_T)}{c_T^{(j-1)}} w_{mi}^{(j-1)} \alpha_{T-1,in}^{(j-1)}.
\end{aligned}
\tag{3.68}
$$

Therefore,

$$
q^{j}_{T,m,i,n}(z_T, z_{T-1}, PE^{*}) = \frac{\phi_{mn}^{(j-1)}(y_T)}{c_T^{(j-1)}} w_{mi}^{(j-1)} \alpha_{T-1,in}^{(j-1)}. \tag{3.69}
$$

In step 2, for $t \in \{2, ..., T-1\}$,we calculate $q^{j}_{t,m,i,n}(z_t, z_{t-1}, PE^{*})$.

$$
\begin{aligned}
&q^j_{t,m,i,n}(z_t, z_{t-1}, PE^*)\\
=&p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^T, \theta^{(j-1)})\\
=&p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^t, y_{t+1}^T, \theta^{(j-1)})\\
=&\frac{p(y_{t+1}^T|z_t = a_m, z_{t-1} = a_i, PE^* = b_n, \theta^{(j-1)})p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^t, \theta^{(j-1)})}{p(y_{t+1}^T|y_1^t, \theta^{(j-1)})}\\
=&\beta^{(j-1)}_{tmn} p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^t, \theta^{(j-1)})\\
=&\beta^{(j-1)}_{tmn} p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^{t-1}, y_t, \theta^{(j-1)})\\
=&\beta^{(j-1)}_{tmn} \frac{p(y_t|z_t = a_m, z_{t-1} = a_i, PE^* = b_n, y_1^{t-1}, \theta^{(j-1)})}{p(y_t|y_1^{t-1}, \theta^{(j-1)})}\\
&p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^{t-1}, \theta^{(j-1)})\\
=&\beta^{(j-1)}_{tmn} \frac{p(y_t|z_t = a_m, PE^* = b_n, \theta^{(j-1)})p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^{t-1}, \theta^{(j-1)})}{p(y_t|y_1^{t-1}, \theta^{(j-1)})}\\
=&\beta^{(j-1)}_{tmn} \frac{\phi^{(j-1)}_{mn}(y_t)}{c_t^{(j-1)}} p(z_t = a_m, z_{t-1} = a_i, PE^* = b_n|y_1^{t-1}, \theta^{(j-1)})\\
=&\beta^{(j-1)}_{tmn} \frac{\phi^{(j-1)}_{mn}(y_t)}{c_t^{(j-1)}} p(z_t = a_m|z_{t-1} = a_i, PE^* = b_n, y_1^{t-1}, \theta^{(j-1)})\\
&p(z_{t-1} = a_i, PE^* = b_n|y_1^{t-1}, \theta^{(j-1)})\\
=&\beta^{(j-1)}_{tmn} \frac{\phi^{(j-1)}_{mn}(y_t)}{c_t^{(j-1)}} p(z_t = a_m|z_{t-1} = a_i, \theta^{(j-1)})p(z_{t-1} = a_i, PE^* = b_n|y_1^{t-1}, \theta^{(j-1)})\\
=&\beta^{(j-1)}_{tmn} \frac{\phi^{(j-1)}_{mn}(y_t)}{c_t^{(j-1)}} w^{(j-1)}_{mi} \alpha^{(j-1)}_{t-1,in}.
\end{aligned}
\tag{3.70}
$$

Therefore, for $t \in \{2, ..., T-1\}$,

$$
q^j_{t,m,i,n}(z_t, z_{t-1}, PE^*) = \beta^{(j-1)}_{tmn} \frac{\phi^{(j-1)}_{mn}(y_t)}{c_t^{(j-1)}} w^{(j-1)}_{mi} \alpha^{(j-1)}_{t-1,in}. \tag{3.71}
$$

Thus, by Eq.(3.67), Eq.(3.69) and Eq.(3.71), the second term is

$$\begin{aligned}
&\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\sum_{t'=2}^{T}\ln p(z_{t'}|z_{t'-1},\theta)\\
&=\sum_{t'=2}^{T-1}\sum_{i,m=1}^{M}\sum_{n=1}^{N}\beta_{tmn}^{(j-1)}\frac{\phi_{mn}^{(j-1)}(y_t)}{c_t^{(j-1)}}w_{mi}^{(j-1)}\alpha_{t-1,in}^{(j-1)}\ln w_{mi}\\
&+\sum_{i,m=1}^{M}\sum_{n=1}^{N}\frac{\phi_{mn}^{(j-1)}(y_T)}{c_T^{(j-1)}}w_{mi}^{(j-1)}\alpha_{T-1,in}^{(j-1)}\ln w_{mi}.
\end{aligned}\tag{3.72}$$

The third term is $\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(z_1|\theta)$.

For all $m \in \{1,...,M\}, n \in \{1,...,N\}$,

$$\begin{aligned}
&\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(z_1|\theta)\\
&=\sum_{z_1}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(z_1|\theta)\\
&=\sum_{m=1}^{M}\sum_{n=1}^{N} q^{(j)}(z_1=a_m, PE^*=b_n)\ln p(z_1=a_m|\theta)\\
&=\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*)\ln u_m\\
&=\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)}\ln u_m.
\end{aligned}\tag{3.73}$$

Thus, for all $m \in \{1,...,M\}, n \in \{1,...,N\}$, the third term is

$$\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(z_1|\theta) = \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)}\ln u_m. \tag{3.74}$$

Similarly, we can calculate the fourth term. For all $m \in \{1,...,M\}, n \in \{1,...,N\}$, the forth term is

$$\sum_{z_1^T}\sum_{PE^*} q^j(z_1^T, PE^*)\ln p(PE^*|\theta) = \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)}\ln v_n. \tag{3.75}$$

Hence, by Eq.(3.60), Eq.(3.67) and Eq.(3.73),

$$
\begin{aligned}
&Q(\theta;\theta^{(j)})\\
&=\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)\ln\phi_{mn}(y_t)+\sum_{t'=2}^{T}\sum_{i,m=1}^{M}\sum_{n=1}^{N} q_{t',m,i,n}^{(j)}(z_{t'}, z_{t'-1}, PE^*)\\
&\quad \ln w_{mi}+\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*)(\ln u_m+\ln v_n).
\end{aligned}
\tag{3.76}
$$

and by Eq.(3.63),Eq.(3.72), Eq.(3.74)and Eq.(3.75),

$$
\begin{aligned}
&Q(\theta;\theta^{(j)})\\
&=\sum_{t=1}^{T-1}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)}\ln\phi_{mn}(y_t)+\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{Tmn}^{(j-1)}\ln\phi_{mn}(y_T)\\
&+\sum_{t'=2}^{T-1}\sum_{i,m=1}^{M}\sum_{n=1}^{N}\beta_{tmn}^{(j-1)}\frac{\phi_{mn}^{(j-1)}(y_t)}{c_t^{(j-1)}}w_{mi}^{(j-1)}\alpha_{t-1,in}^{(j-1)}\ln w_{mi}\\
&+\sum_{i,m=1}^{M}\sum_{n=1}^{N}\frac{\phi_{mn}^{(j-1)}(y_T)}{c_T^{(j-1)}}w_{mi}^{(j-1)}\alpha_{T-1,in}^{(j-1)}\ln w_{mi}+\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{1mn}^{(j-1)\beta_{1mn}^{(j-1)}}(\ln u_m+\ln v_n).
\end{aligned}
$$

Experts can put their knowledge into the parameter estimation procedure via $p(\theta)$ in Eq.(3.51). Here, we assume that all parameters are independent, $p(\theta) = p(\sigma)p(\mathbf{u})p(\mathbf{v})p(\mathbf{W})$. In our experience, investment experts usually have two types of knowledge which are useful to estimate θ. The first type of knowledge is about PE^*. Often, experts may be able to estimate the range of appropriate PE^* level by analyzing a firm's business strategy together with competitions in its industry. The second type of knowledge is about the degree of persistence of the medium-term noisy effect which makes a stock price deviate from its fundamental for a considerable amount of time as explained in Section 3.2. For some firms, e.g. a firm with non-existent investor relation department, when there exist some unconfirmed rumors, its price can deviate from its fundamental for a long period. In contrast, some firms with both strong public and investor relation departments can clear up unconfirmed rumors rather quickly, so this rumor effect will not stay long. The two types of expert information can be encoded on $p(\mathbf{v})$ and $p(\mathbf{W})$, respectively. The prior $p(\mathbf{v})$ for the vector $\mathbf{v} = (v_n)_{N\times 1}$ can be represented via the Dirichlet distribution:

$$p(\mathbf{v}) = \frac{\tau(k_1 + k_2 + ... + k_N)}{\tau(k_1)\tau(k_2)...\tau(k_N)} \prod_{n=1}^{N} v_n^{k_n - 1} \tag{3.77}$$

Intuitively, $k_n, n \in \{1, ..., N\}$ is a degree of belief for each possible PE^* value b_n. Experts can employ their believes that some value of PE^*, e.g. b_i is relatively more probable than other values by giving k_i relatively higher value than other $k_n, n \neq i$. See (Gelman et al., 2003) for more details on the Dirichlet prior. The prior on transition matrix $p(\mathbf{W})$ encoding the average persistence degree of the medium-term noisy effect can also be described by a product of Dirichlet priors (Strelioff et al., 2007) : $p(W) = \prod_{m=1}^{M} p(\mathbf{w}_m)$, where as defined in Section 3.2, $\mathbf{w}_m = (w_{im})_{i=1,...,M}$ denotes a $M \times 1$ vector of a probability $p(z_{t+1} = a_i | z_t = a_m), i = 1, ..., M$, and

$$p(\mathbf{w}_m) = \frac{\tau(\sum_{i=1}^{M} k_{im})}{\prod_{i=1}^{M} \tau(k_{im})} \prod_{i=1}^{M} w_{im}^{k_{im}-1}. \tag{3.78}$$

The persistence degree of the medium-term effect can be set by relatively increasing the values of k_{mm} compared to other values $k_{im}, i \neq m$. The relatively higher of k_{mm}, the more persistence of the medium-term effect. Since medium-term effect appears at random, other values can symmetrically be set: $k_{im} = (1 - k_{mm})/(M - 1)$, for $i \neq m$.

Therefore, the log prior probability $\ln p(\theta)$ is given by

$$\begin{aligned}
&\ln p(\theta) \\
&= \ln p(\mathbf{W}) + \ln p(\mathbf{v}) + constant \\
&= \sum_{m=1}^{M}\sum_{i=1}^{M}(k_{im}-1)\ln w_{im} + \sum_{n=1}^{N}(k_n - 1)\ln v_n + constant.
\end{aligned} \tag{3.79}$$

Combining with the $\ln p(\theta)$ term described below, the constraint maximization Eq.(3.52) is well defined and readily to be solved by using the method of *Lagrange multipliers* (S. Boyd and Hassibi, 2007). We have two types of constraints which are equality constraints and inequality constraints. These equality constraints are $\sum_{m=1}^{M} u_m = 1$,$\sum_{n=1}^{N} v_n = 1$,and $\sum_{m=1}^{M} w_{im} = 1$. The inequality constraints are $0 \le u_m \le 1$,$0 \le v_n \le 1$,$0 \le w_{im} \le 1$ and $\sigma > 0$. According to the equality constraints, by Eq.(3.76)and Eq.(3.79), the Lagrange function is given by

$$\begin{aligned}
&f(\theta, \lambda_1, \lambda_2, \lambda_3) \\
=&Q(\theta;\theta^{(j)}) + \ln p(\theta) + \lambda_1(\sum_{m'} u_{m'} - 1) + \lambda_2(\sum_{n'} v_{n'} - 1) + \lambda_3(\sum_{m'} w_{m'i} - 1) \\
&+ constant \\
=&\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)\ln\phi_{mn}(y_t) + \sum_{t'=2}^{T}\sum_{i,m=1}^{M}\sum_{n=1}^{N} q_{t',m,i,n}^{(j)}(z_{t'}, z_{t'-1}, PE^*)\ln w_{mi} \\
&+ \sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*)(\ln u_m + \ln v_n) + \sum_{m=1}^{M}\sum_{i=1}^{M}[k_{im}-1]\ln w_{im} + \sum_{n=1}^{N}(k_n - 1) \\
&\ln v_n + \lambda_1(\sum_{m'} u_{m'} - 1) + \lambda_2(\sum_{n'} v_{n'} - 1) + \lambda_3(\sum_{m'} w_{m'i} - 1) + constant.
\end{aligned} \tag{3.80}$$

3.3.2.2 M-step calculation

Next, we calculate the partial derivative related to these four unknown parameters which are $u_m, v_n, w_{mi}, \sigma^2$ and then check these parameters satisfy the inequality constraints.

(a) Calculate $u_m^{(j)}, j = 2, 3, ...$

For each fixed $m \in \{1, ..., M\}$and $j \in \{2, 3, ...\}$, let $\dfrac{\partial f}{\partial u_m} = 0$.

$$\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*)\frac{1}{u_m} + \lambda_1 = 0$$
$$\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*) + \lambda_1 u_m = 0$$
$$\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*) + \sum_{m=1}^{M} \lambda_1 u_m = 0$$
$$\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*) + \lambda_1 \sum_{m=1}^{M} u_m = 0$$
$$\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*) + \lambda_1 = 0$$
$$\lambda_1 = -\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*) = -1. \tag{3.81}$$

Then,

$$u_m^{(j)} = \sum_{n=1}^{N} \alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)}. \tag{3.82}$$

and $0 \le u_m^{(j)} \le 1$ satisfies the inequality constraint $0 \le u_m \le 1$.

Therefore, for each fixed $m \in \{1, ..., M\}$and $j \in \{2, 3, ...\}$,

$$u_m^{(j)} = \sum_{n=1}^{N} \alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)}.$$

(b) Calculate $v_n^{(j)}, j = 2, 3, ...$

For each fixed $n \in \{1, ..., N\}$ and $j \in \{2, 3, ...\}$, let $\dfrac{\partial f}{\partial v_n} = 0$.

$$
\begin{aligned}
&\sum_{m=1}^{M} q_{1mn}^{(j)}(z_1, PE^*)\frac{1}{v_n} + (k_n - 1)\frac{1}{v_n} + \lambda_2 = 0 \\
&\sum_{m=1}^{M} q_{1mn}^{(j)}(z_1, PE^*) + (k_n - 1) + \lambda_2 v_n = 0 \\
&\sum_{m=1}^{M}\sum_{n=1}^{N} q_{1mn}^{(j)}(z_1, PE^*) + \sum_{n=1}^{N}(k_n - 1) + \sum_{n=1}^{N}\lambda_2 v_n = 0 \\
&1 + \sum_{n=1}^{N}(k_n - 1) + \lambda_2 = 0 \\
&\lambda_2 = -1 - \sum_{n=1}^{N}(k_n - 1). \qquad (3.83)
\end{aligned}
$$

Then,

$$
\begin{aligned}
&v_n^{(j)} \\
&= \frac{\sum\limits_{m=1}^{M} q_{1mn}^{(j)}(z_1, PE^*) + (k_n - 1)}{1 + \sum\limits_{n'=1}^{N}(k_{n'} - 1)} \\
&= \frac{\sum\limits_{m=1}^{M} \alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)} + (k_n - 1)}{1 + \sum\limits_{n'=1}^{N}(k_{n'} - 1)}, \qquad (3.84)
\end{aligned}
$$

and $0 \leq v_n^{(j)} \leq 1$ satisfies the inequality constraint $0 \leq v_n \leq 1$.

Therefore, for each fixed $n \in \{1, ..., N\}$and $j \in \{2, 3, ...\}$,

$$
v_n^{(j)} = \frac{\sum\limits_{m=1}^{M} \alpha_{1mn}^{(j-1)}\beta_{1mn}^{(j-1)} + (k_n - 1)}{1 + \sum\limits_{n'=1}^{N}(k_{n'} - 1)}. \qquad (3.85)
$$

(c) Calculate $w_{mi}^{(j)}, j = 2, 3, ...$

For all fixed $m, i \in \{1, ..., M\}$,and $j \in \{2, 3, ...\}$, let $\dfrac{\partial f}{\partial w_{mi}} = 0$.

$$\sum_{t'=2}^{T}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*})\frac{1}{w_{mi}} + (k_{mi}-1)\frac{1}{w_{mi}} + \lambda_3 = 0$$

$$\sum_{t'=2}^{T}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) + (k_{mi}-1) + \lambda_3 w_{mi} = 0$$

$$\sum_{t'=2}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) + \sum_{m=1}^{M}(k_{mi}-1) + \sum_{m=1}^{M}\lambda_3 w_{mi} = 0$$

$$\sum_{t'=2}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) + \sum_{m=1}^{M}(k_{mi}-1) + \lambda_3 \sum_{m=1}^{M} w_{mi} = 0$$

$$\sum_{t'=2}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) + \sum_{m=1}^{M}(k_{mi}-1) + \lambda_3 = 0$$

$$\lambda_3 = -\sum_{t'=2}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) - \sum_{m=1}^{M}(k_{mi}-1). \tag{3.86}$$

Then,

$$\begin{aligned} & w^{(j)}_{mi} \\ &= -\frac{\sum_{t'=2}^{T}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) + (k_{mi}-1)}{\sum_{t'=2}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q^{(j)}_{t',m,i,n}(z_{t'}, z_{t'-1}, PE^{*}) + \sum_{m'=1}^{M}(k_{m'i}-1)} \\ &= \frac{mol}{den}. \end{aligned} \tag{3.87}$$

where

$$mol = \sum_{t'=2}^{T-1}\sum_{n=1}^{N} \beta^{(j-1)}_{t'mn}\frac{\phi^{(j-1)}_{mn}(y_{t'})}{c^{(j-1)}_{t'}} w^{(j-1)}_{mi}\alpha^{(j-1)}_{t'-1,i,n} + \sum_{n=1}^{N}\frac{\phi^{(j-1)}_{mn}(y_T)}{c^{(j-1)}_{T}} w^{(j-1)}_{mi}\alpha^{(j-1)}_{T-1,i,n} + (k_{mi}-1), \tag{3.88}$$

and

$$den = \sum_{m'=1}^{M} molecular. \tag{3.89}$$

Meanwhile, $0 \le w^{(j)}_{mi} \le 1$ satisfies the inequality constraint $0 \le w_{mi} \le 1$.

Therefore, for all fixed $m, i \in \{1, ..., M\}$,and $j \in \{2, 3, ...\}$,

$$w^{(j-1)}_{mi} = \frac{mol}{den}. \tag{3.90}$$

(d) Calculate $(\sigma^2)^{(j)}, j = 2, 3, ...$

For all $m \in \{1, ..., M\}$,and $n \in \{1, ..., N\}$,

$$\phi(y_t) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(y_t - \ln(b_n(1 + a_m)))^2}{2\sigma^2}). \tag{3.91}$$

$$\begin{aligned}
&\frac{\partial\phi_{mn}(y_t)}{\partial\sigma}\\
&= -\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma^2}\exp(-\frac{(y_t - \ln(b_n(1 + a_m)))^2}{2\sigma^2})\\
&+\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(y_t - \ln(b_n(1 + a_m)))^2}{2\sigma^2})(y_t - \ln(b_n(1 + a_m))^2)\frac{1}{\sigma^3}\\
&=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(y_t - \ln(b_n(1 + a_m)))^2}{2\sigma^2})(-\frac{1}{\sigma} + \frac{(y_t - \ln(b_n(1 + a_m)))^2}{\sigma^3})\\
&=\phi_{mn}(y_t)(-\frac{1}{\sigma} + \frac{(y_t - \ln(b_n(1 + a_m)))^2}{\sigma^3}).
\end{aligned} \tag{3.92}$$

Let $\frac{\partial f}{\partial\sigma} = 0$, then

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)\frac{1}{\phi_{mn}(y_t)}\phi_{mn}(y_t)(-\frac{1}{\sigma} + \frac{(y_t - \ln(b_n(1 + a_m)))^2}{\sigma^3}) = 0$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)(-\frac{1}{\sigma} + \frac{(y_t - \ln(b_n(1 + a_m)))^2}{\sigma^3}) = 0$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)(-\sigma^2 + (y_t - \ln(b_n(1 + a_m)))^2) = 0$$

$$\begin{aligned}
\sigma^2 &= \frac{\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)(y_t - \ln(b_n(1 + a_m)))^2}{\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N} q_{tmn}^{(j)}(z_t, PE^*)}\\
&=\frac{\sum_{t=1}^{T-1}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)}(y_t - \ln(b_n(1 + a_m)))^2}{\sum_{t=1}^{T-1}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)} + \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{Tmn}^{(j-1)}}\\
&+\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{Tmn}^{(j-1)}(y_T - \ln(b_n(1 + a_m)))^2}{\sum_{t=1}^{T-1}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)} + \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{Tmn}^{(j-1)}}.
\end{aligned} \tag{3.93}$$

Therefore,

$$
\begin{aligned}
&(\sigma^2)^{(j-1)}\\
&=\frac{\sum\limits_{t=1}^{T-1}\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)}(y_t-\ln(b_n(1+a_m)))^2}{\sum\limits_{t=1}^{T-1}\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)}+\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N}\alpha_{Tmn}^{(j-1)}}\\
&+\frac{\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N}\alpha_{Tmn}^{(j-1)}(y_T-\ln(b_n(1+a_m)))^2}{\sum\limits_{t=1}^{T-1}\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N}\alpha_{tmn}^{(j-1)}\beta_{tmn}^{(j-1)}+\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N}\alpha_{Tmn}^{(j-1)}}.
\end{aligned}
\tag{3.94}
$$

Note that, for all $m, i \in \{1, ..., M\}$, Eq.(3.90) is complicated and

$$
\sum_{t'=2}^{T-1}\sum_{n=1}^{N}\beta_{t'mn}^{(j-1)}\frac{\phi_{mn}^{(j-1)}(y_{t'})}{c_{t'}^{(j-1)}}w_{mi}^{(j-1)}\alpha_{t'-1,i,n}^{(j-1)}+\sum_{n=1}^{N}\frac{\phi_{mn}^{(j-1)}(y_T)}{c_T^{(j-1)}}w_{mi}^{(j-1)}\alpha_{T-1,i,n}^{(j-1)}, \tag{3.95}
$$

a part of $w_{mi}^{(j)}$ in Eq.(3.90) and we write it in a matrix form below, so that we can write a computer program. Let $\widetilde{\mathbf{W}^j} = (\widetilde{w_{mn}^j})_{M\times N}$.

$$
\begin{aligned}
&\begin{pmatrix}
\widetilde{w_{11}^j} & \widetilde{w_{12}^j} & \cdots & \widetilde{w_{1M}^j}\\
\widetilde{w_{21}^j} & \widetilde{w_{22}^j} & \cdots & \widetilde{w_{2M}^j}\\
\cdots & \cdots & \cdots & \cdots\\
\widetilde{w_{M1}^j} & \widetilde{w_{M2}^j} & \cdots & \widetilde{w_{MM}^j}
\end{pmatrix}
=
\begin{pmatrix}
w_{11}^{j-1} & w_{12}^{j-1} & \cdots & w_{1M}^{j-1}\\
w_{21}^{j-1} & w_{22}^{j-1} & \cdots & w_{2M}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
w_{M1}^{j-1} & w_{M2}^{j-1} & \cdots & w_{MM}^{j-1}
\end{pmatrix}\circ\\
&\left[\frac{1}{c_2^{j-1}}
\begin{pmatrix}
\beta_{211}^{j-1} & \beta_{212}^{j-1} & \cdots & \beta_{21N}^{j-1}\\
\beta_{221}^{j-1} & \beta_{222}^{j-1} & \cdots & \beta_{22N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\beta_{2M1}^{j-1} & \beta_{2M2}^{j-1} & \cdots & \beta_{2MN}^{j-1}
\end{pmatrix}\circ
\begin{pmatrix}
\phi_{11}^{j-1}(y_2) & \phi_{12}^{j-1}(y_2) & \cdots & \phi_{1N}^{j-1}(y_2)\\
\phi_{21}^{j-1}(y_2) & \phi_{22}^{j-1}(y_2) & \cdots & \phi_{2N}^{j-1}(y_2)\\
\cdots & \cdots & \cdots & \cdots\\
\phi_{M1}^{j-1}(y_2) & \phi_{M2}^{j-1}(y_2) & \cdots & \phi_{MN}^{j-1}(y_2)
\end{pmatrix}\circ
\begin{pmatrix}
\alpha_{111}^{j-1} & \alpha_{112}^{j-1} & \cdots & \alpha_{11N}^{j-1}\\
\alpha_{121}^{j-1} & \alpha_{122}^{j-1} & \cdots & \alpha_{12N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\alpha_{1M1}^{j-1} & \alpha_{1M2}^{j-1} & \cdots & \alpha_{1MN}^{j-1}
\end{pmatrix}^T\right.\\
&+\frac{1}{c_3^{j-1}}
\begin{pmatrix}
\beta_{311}^{j-1} & \beta_{312}^{j-1} & \cdots & \beta_{31N}^{j-1}\\
\beta_{321}^{j-1} & \beta_{322}^{j-1} & \cdots & \beta_{32N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\beta_{3M1}^{j-1} & \beta_{3M2}^{j-1} & \cdots & \beta_{3MN}^{j-1}
\end{pmatrix}\circ
\begin{pmatrix}
\phi_{11}^{j-1}(y_3) & \phi_{12}^{j-1}(y_3) & \cdots & \phi_{1N}^{j-1}(y_3)\\
\phi_{21}^{j-1}(y_3) & \phi_{22}^{j-1}(y_3) & \cdots & \phi_{2N}^{j-1}(y_3)\\
\cdots & \cdots & \cdots & \cdots\\
\phi_{M1}^{j-1}(y_3) & \phi_{M2}^{j-1}(y_3) & \cdots & \phi_{MN}^{j-1}(y_3)
\end{pmatrix}\circ
\begin{pmatrix}
\alpha_{211}^{j-1} & \alpha_{212}^{j-1} & \cdots & \alpha_{21N}^{j-1}\\
\alpha_{221}^{j-1} & \alpha_{222}^{j-1} & \cdots & \alpha_{22N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\alpha_{2M1}^{j-1} & \alpha_{2M2}^{j-1} & \cdots & \alpha_{2MN}^{j-1}
\end{pmatrix}^T+\cdots\\
&+\frac{1}{c_{T-1}^{j-1}}
\begin{pmatrix}
\beta_{T-1,11}^{j-1} & \beta_{T-1,12}^{j-1} & \cdots & \beta_{T-1,1N}^{j-1}\\
\beta_{T-1,21}^{j-1} & \beta_{T-1,22}^{j-1} & \cdots & \beta_{T-1,2N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\beta_{T-1,M1}^{j-1} & \beta_{T-1,M2}^{j-1} & \cdots & \beta_{T-1,MN}^{j-1}
\end{pmatrix}\circ
\begin{pmatrix}
\phi_{11}^{j-1}(y_{T-1}) & \phi_{12}^{j-1}(y_{T-1}) & \cdots & \phi_{1N}^{j-1}(y_{T-1})\\
\phi_{21}^{j-1}(y_{T-1}) & \phi_{22}^{j-1}(y_{T-1}) & \cdots & \phi_{2N}^{j-1}(y_{T-1})\\
\cdots & \cdots & \cdots & \cdots\\
\phi_{M1}^{j-1}(y_{T-1}) & \phi_{M2}^{j-1}(y_{T-1}) & \cdots & \phi_{MN}^{j-1}(y_{T-1})
\end{pmatrix}\circ
\begin{pmatrix}
\alpha_{T-2,11}^{j-1} & \alpha_{T-2,12}^{j-1} & \cdots & \alpha_{T-2,1N}^{j-1}\\
\alpha_{T-2,21}^{j-1} & \alpha_{T-2,22}^{j-1} & \cdots & \alpha_{T-2,2N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\alpha_{T-2,M1}^{j-1} & \alpha_{T-2,M2}^{j-1} & \cdots & \alpha_{T-2,MN}^{j-1}
\end{pmatrix}^T\\
&\left.+\frac{1}{c_T^{j-1}}
\begin{pmatrix}
\phi_{11}^{j-1}(y_T) & \phi_{12}^{j-1}(y_T) & \cdots & \phi_{1N}^{j-1}(y_T)\\
\phi_{21}^{j-1}(y_T) & \phi_{22}^{j-1}(y_T) & \cdots & \phi_{2N}^{j-1}(y_T)\\
\cdots & \cdots & \cdots & \cdots\\
\phi_{M1}^{j-1}(y_T) & \phi_{M2}^{j-1}(y_T) & \cdots & \phi_{MN}^{j-1}(y_T)
\end{pmatrix}\circ
\begin{pmatrix}
\alpha_{T-1,11}^{j-1} & \alpha_{T-1,12}^{j-1} & \cdots & \alpha_{T-1,1N}^{j-1}\\
\alpha_{T-1,21}^{j-1} & \alpha_{T-1,22}^{j-1} & \cdots & \alpha_{T-1,2N}^{j-1}\\
\cdots & \cdots & \cdots & \cdots\\
\alpha_{T-1,M1}^{j-1} & \alpha_{T-1,M2}^{j-1} & \cdots & \alpha_{T-1,MN}^{j-1}
\end{pmatrix}^T\right]
\end{aligned}
\tag{3.96}
$$

Therefore,

$$
\begin{aligned}
&\widetilde{\mathbf{W}^{(j)}} \\
=&\mathbf{W}^{(j-1)} \circ \left[\frac{1}{c_2^{j-1}} \mathbf{B}_2^{(j-1)} \circ \mathbf{\Phi}_2^{(j-1)} \mathbf{A}_1^{(j-1)T} + \frac{1}{c_3^{j-1}} \mathbf{B}_3^{(j-1)} \circ \mathbf{\Phi}_3^{(j-1)} \mathbf{A}_2^{(j-1)T} \right] + ... \\
+&\mathbf{W}^{(j-1)} \circ \left[\frac{1}{c_{T-1}^{j-1}} \mathbf{B}_{T-1}^{(j-1)} \circ \mathbf{\Phi}_{T-1}^{(j-1)} \mathbf{A}_{T-1}^{(j-1)T} + \frac{1}{c_T^{j-1}} \mathbf{\Phi}_T^{(j-1)} \mathbf{A}_{T-1}^{(j-1)T} \right]. \qquad (3.97)
\end{aligned}
$$

CHAPTER IV

EXPERIMENT BASED ON PE RATIO

In this section, we illustrate benefits of our methodology in real-world applications. To do this, we will conduct comprehensive trading simulations to show consistent superior performances of our method over standard benchmark. In literatures of finance, according to the Efficient Market Hypothesis (Fama, 1976, 1991; Mark, 2011) which is widely accepted by mainstream researchers, the gold standard benchmark is, surprisingly, the simple buy-and-hold method which empirically proves to be efficient in the long run. Astonishingly, much evidence clearly indicates that most mutual fund managers who apply complex active portfolio management techniques cannot beat the simple buy-and-hold strategy of the market portfolio (Malkiel, 2016; Carhart, 1997). In this section, we will test our method against this gold standard buy-and-hold strategy both on individual stock level and on portfolio level.

4.1 The data

We collected the data sets from SET in Thailand, NYSE and NASDAQ in US. In the thesis, the data sets are daily stock prices, quarterly earnings to calculate the yearly earning. The data sets are over a 5-year period from Jan 1, 2012 to Sep 30, 2016. Public holidays, Saturday and Sunday prices were removed to avoid any bias in the results from weekend market closures. The historical prices and the historical earnings are adjusted according to stock splits. The criterion for our model is that the historical earnings are positive. Most firms satisfy this criterion. All data are adjusted for stock splitting if occuring during this 5-year period.

In this section, we will make stock trading in the markets of two different countries where we can access historical data: NYSE (New York Stock Exchange) and NASDAQ in US and SET (Stock Exchange of Thailand) in Thailand. While NYSE and NASDAQ represent mature stock markets, SET represents an emerging market, so that we are able to test our methodology to firms in both market phases. For each country, we collected data of 10 firms from various industries to ensure that our methodology is not just limited to one specific industry. Each selected firm is well established and has at least 5 year historical trading data. The names of selected companies with their respective sectors for Thai stocks and US stocks are shown in Table 4.1 and Table 4.2, respectively.

We collected daily 5-year historical closing-price data for each firm from Jan 1, 2012 to Sep 30, 2016 consisting of 1160 closing prices for stocks in SET and 1195 closing prices for stocks in NYSE and NASDAQ, respectively. The difference in the number of data is due to different working days in the two countries.

4.2 Experiment setting

To avoid duplicated writing, here, we shall explain only experiment settings for stocks in SET with historical price $P_1, ..., P_{1160}$. The experiment settings for stocks in NYSE and NASDAQ are done similarly.

For each firm, the corresponding yearly earnings data in those years are also collected. $E_1, ..., E_{1160}$ are defined as the summation of the most recent 4 quarterly earnings on each data t. The first 3-year historical data (Jan 1,2012 to Dec 31,2014) $P_1, ..., P_{735}$ and $E_1, ..., E_{735}$ will be used as a training data for our Bayesian methodology to learn the appropriate parameters $\theta = \{\mathbf{W},\mathbf{u},\mathbf{v}, \sigma^2\}$

Table 4.1 Stock symbols of selected firms from Thailand, including their industries. BigCap, MidCap and SmallCap are defined according to market sizes (in THB) which are greater than 100 billions, 10 billions and 1 billions, respectively.

Symbol	Industry	Size
CPALL	Retailing-Food and Staples	BigCap
CPN	Developer-Department Stores	BigCap
EASTW	Utilities-Water Resources	MidCap
GLOW	Utilities-Power Plant	BigCap
HMPRO	Retailing-Household Products	BigCap
QH	Developer-Housing	MidCap
ROBINS	Retailing-General	MidCap
SCB	Banking	BigCap
SNC	Electrical Equipments	SamllCap
TTW	Utilities-Tap Water	MidCap

using the EM algorithm as well as estimate the most probable values of PE^* and $\{z_1, ..., z_{735}\}$ by the method of smoothing as explained in Section 3.3. The constants $\{a_1, ..., a_M\}$ and $\{b_1, ..., b_N\}$ are set by experts. Since the constant M determines the size of the transition matrix $\mathbf{W} = (w_{im})_{M\times M}$, we make a constraint $M < 10$ so that the model is not over-parameterized and that 3-year historical data is enough to learn $\mathbf{W}$. For all prior distributions, we employ non-informative priors with the exception of $p(\mathbf{W})$ where our "security experts" emphasize the prior knowledge of z_t persistency as described in Subsection 3.3.2.

Each trading simulation is conducted for each individual stock with the remaining 2-year historical data $P_{736}, ..., P_{1160}$ to measure the performance of both

Table 4.2 Stock symbols of selected firms from US, including their industries and market sizes in USD. B denotes billion dollars (Data retrieved on Dec 7, 2016).

Symbol	Industry	Size
WMT	Services-Discount and Variety stores	218.48B
HD	Services-Home improvement stores	157.86B
KO	Consumer Good-Beverages soft drinks	173.51B
G	Services-Business services	4.925B
AAPL	Consumer Goods-Electronic equipment	584.37B
NKE	Consumer Goods-Textile-Apparel Footwear and Accessories	84.35B
BK	Financial-Asset Management	51.89B
CF	Basic Materials-Agricultural Chemicals	6.695B
CSCO	Technology- Networking and Communication Devices	147.88B
DIS	Services-Entertainment Diversified	157.41B

our method and the benchmark. The performance measurement metric is, as used by practitioners, a profit generated by each method. The profit calculation is straightforward: for each trading simulation, each method is equally given an initial amount of cash I to make a trade (which taken into account a commission fee), and the profit are simply all the asset values at the end of a simulation minus I. For simplicity, we assume that each stock can be bought with all the money we have, e.g. supposing we have 100$ and a stock's price is 12$, then we are able to buy $100/12 = 8.33$ stocks.

4.2.1 Buy-and-hold strategy

Buy-and-hold strategy is very simple and it is widely used. (Preis et al., 2013; Enke and Thawornwong, 2005; Chen et al., 2003; Agarwal and Naik, 2004) use this strategy as the benchmark. We use buy-and-hold strategy for comparison in the thesis. This benchmark is to buy the stock with all the cash at the beginning and then do nothing until the end. Initially, this method will get $C.I/P_{736}$ shares where $C \approx 0.9987$ represents the value of assets after taking SET's commission fee into account. At the end of the simulation this asset will have a value of $P_{1160}.C.I/P_{736}$, so the profit can be calculated easily.

4.2.2 Long-term strategy and medium-term strategy

For a trading strategy employed by our method, there are two possible versions inspired by our model's main idea (see Figure 3.1) and Strategy A (buy low, sell high) described in Section 3.1. The first version called long-term strategy is simply to "buy low, sell high" with respect to the static value of PE^*, and the second version called medium-term strategy is to "buy low, sell high" with respect to the dynamic values of $PE^*(1+z_t)$ where each z_t is dynamically estimated by the method of filtering described in Subsection 3.3.2. Both versions can be formally described as follows.

Let I_t and N_t be available cash and total shares at date t, respectively. Initially, $I_{736} = I$ and $N_{736} = 0$. Now, both trading versions can be defined simply by the following procedure: for each date t, exactly one of the following cases holds:

(i) $P_t/E_t \leq A_t(1 - Tr)$ and $I_t > 0$ (buy-low case) where $Tr \in (0, 1)$ is a threshold, $A_t = PE^*$ for the long-term strategy and $A_t = PE^*(1 + z_t)$ for the medium-term strategy. In this case, buy the stock with all cash, so that

$N_{t+1} = C.I_t/P_T$ and $I_{t+1} = 0$.

(ii) $P_t/E_t \geq A_t(1+Tr)$ and $I_t = 0$ (sell-high case). In this case, sell all the holding stock to get cash $I_{t+1} = P_t.N_t.C$ and $N_{t+1} = 0$.

(iii) If case (i) and case (ii) are not satisfied, do nothing. So, $I_{t+1} = I_t$ and $N_{t+1} = N_t$. At the end of a trading simulation $t = 1160$, the total profit is simply $I_{1160} + P_{1160}.N_{1160} - I_{736}$, so that we can compare with the buy-and-hold strategy profit.

We give some illustrations of our trading in actions which are shown in Figure 4.1 and Figure 4.2. Figure 4.1 is an example of long-term trading of CPALL with threshold 0.05 and Figure 4.2 is an example of medium-term trading of CPALL with threshold 0.05. In Figure 4.1 “Green circle” denotes “buy” and “Black cross” denotes “sell”. Top and bottom figures show the same trading in different perspectives The top figure shows trading with respect to the “PE” perspective where Red line denotes PE^*. Here, it is easy to see our strategy in action: when the observed PE is lower or higher than the threshold level, buy or sell is triggered, respectively. The bottom figure shows trading in the “price” perspective when Red line denotes $P^* = E_t PE^*$. Since the earnings continue to increase, P^* also increase accordingly. In Figure 4.2, in addition to those explained in Figure 4.1, in the top figure, the purple dashed line denotes $PE^*(1+z_t)$, and becomes the base line of this trading strategy. Note that our Bayesian method estimates the purple line by the method of “filtering” which tracks the observed PE movements with some delay. The bottom figure is the “price” perspective where the purple line shows $P^*(1+z_t)$. In this example, buy-and-hold strategy beats ours by small margin because of the commission fees caused by our frequent trading.

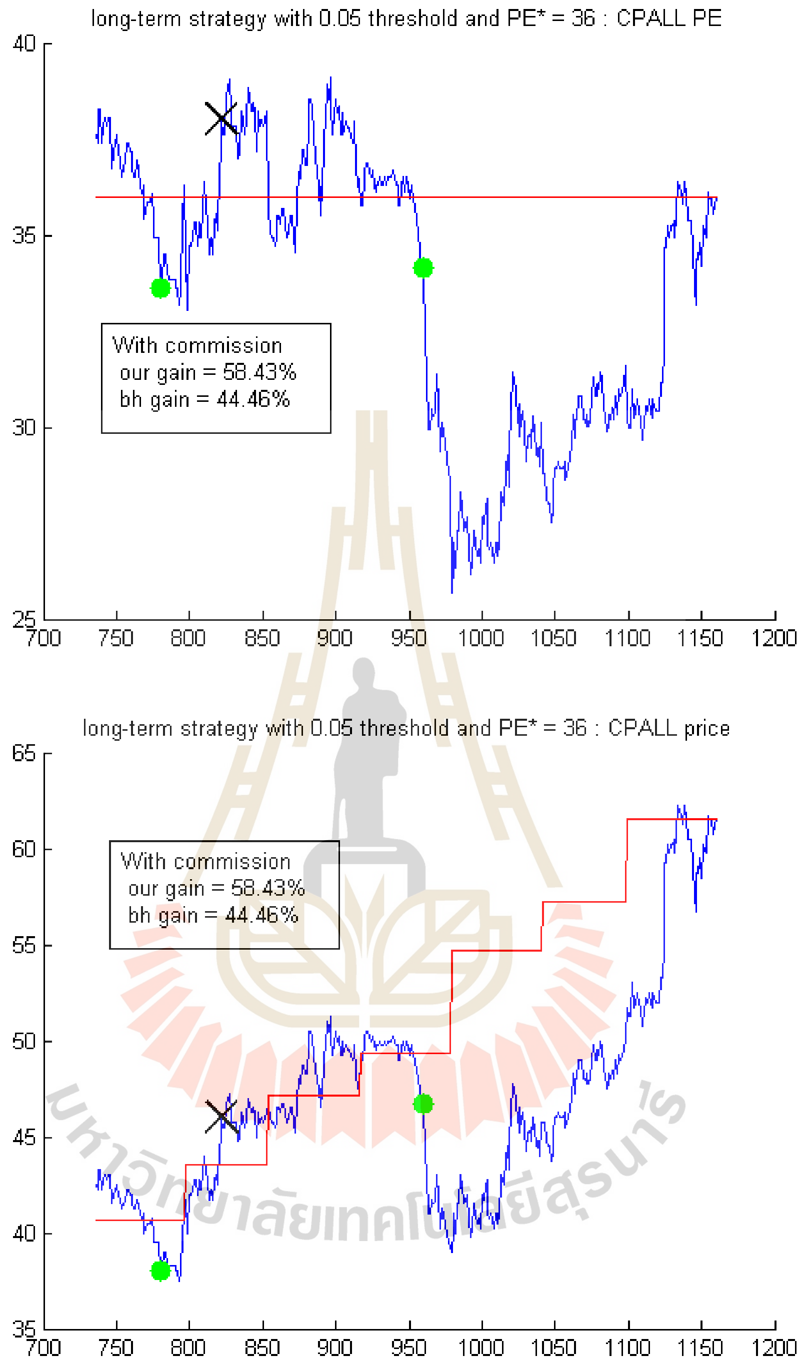

Figure 4.1 Example of Long-term strategy trading of CPALL with threshold 5% where our model's profit is 58.43% while buy-and-hold strategy profit is 44.46%.

4.2.3 The diagram

Before stock trading simulations, the process is seen in Figure 4.3 .

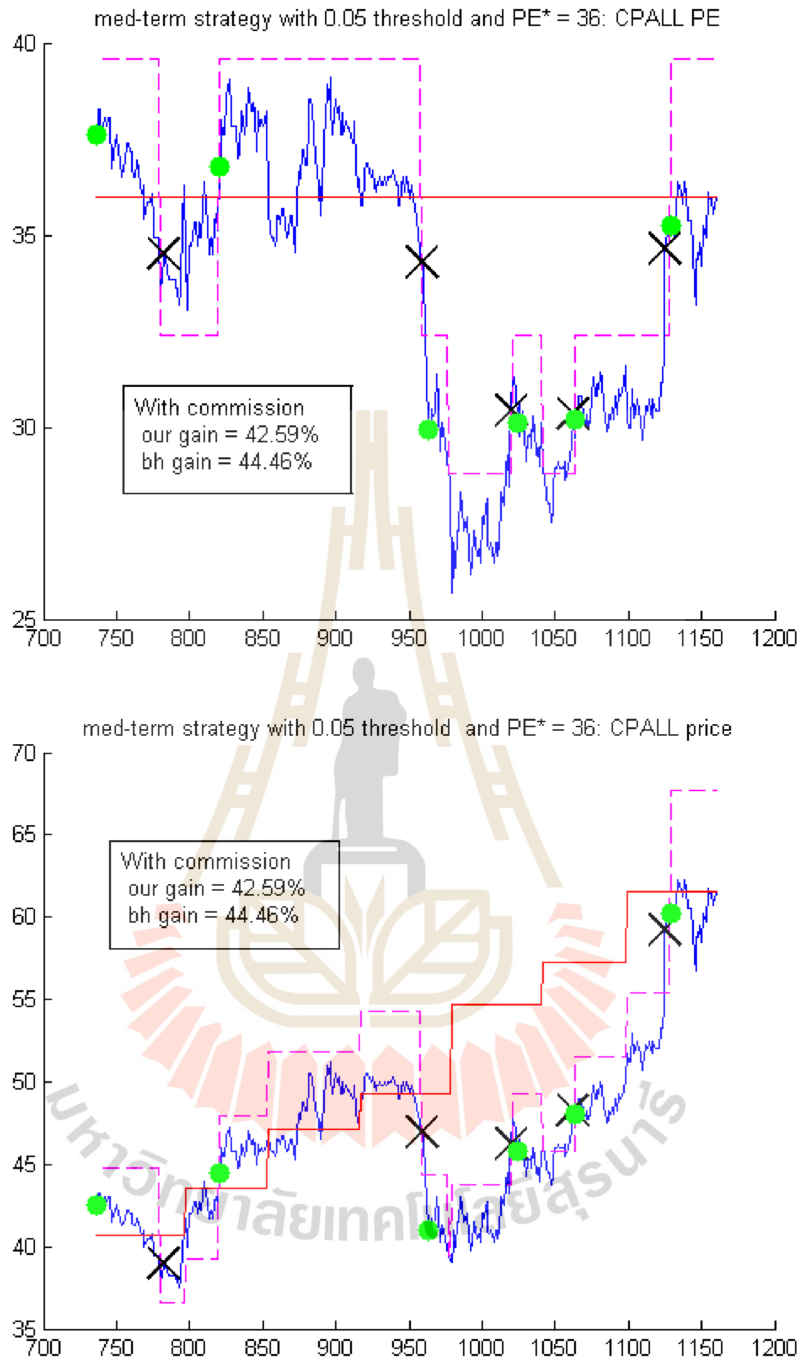

Figure 4.2 Example of Medium-term strategy trading of CPALL with threshold $Tr = 5\%$.

Step 1: Collect the historical data (5-year data) as described in Section 4.2. The first 3-year data is considered as a training data and the following 2-year data

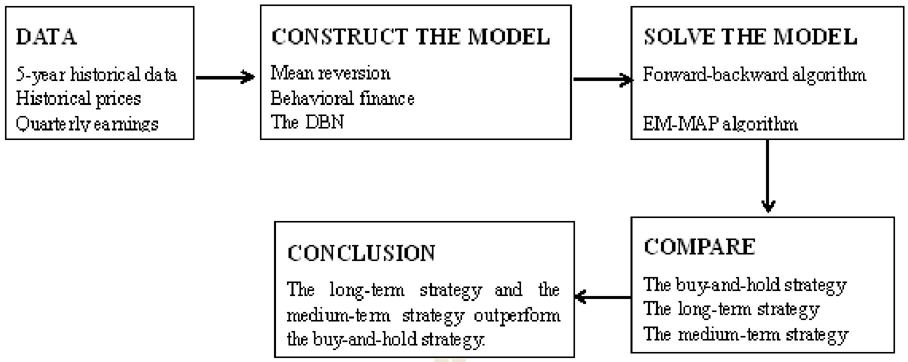

Figure 4.3 The diagram of stock trading based on PE ratio

is considered as a testing data.

Step 2: Based on mean reversion and behavioral volatility, we construct our model combining fundamental valuation with the DBN, see Chapter III.

Step 3: For the training data, calculate the parameters by forward-backward algorithm and EM-MAP algorithm and more details can be found in Section 3.3. For the training data, we use forward-backward algorithm again and estimate latent variables PE^* and the medium-effect z_t by smoothing. For the testing data, we estimate the medium-effect z_t by filtering.

Note that small probability events appear and we modify them before next step. The small probability events are explained in Figure 4.4. In Figure 4.4,the red circle includes small probability events which continue for a very short period. We modify the values of z_t to be the same as the values of z_t before the period.

Step 4: Long-term stock trading strategy and medium-term strategy are given, as described in Section 4.1. We compare those two versions of stock trading strategy with buy-and-hold strategy, respectively.

Step 5: The long-term strategy and the medium-term strategy outperform the buy-and-hold strategy.

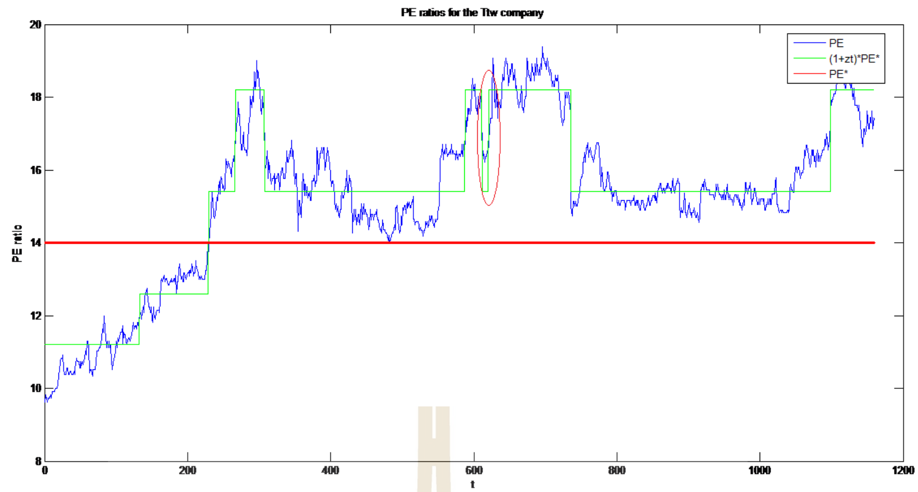

Figure 4.4 Example of small probability events of TTW .

4.3 Experimental results and discussions

4.3.1 Individual firm-level experiments

To ensure that our experimental results are not biased because of a threshold choice, we test 4 different thresholds for each trading strategy. Note that the thresholds in the medium-term trading are relatively smaller than those in the long-term. This is due to the nature of medium-term strategy where PE_t has a smaller deviation from its base line $PE^*(1+z_t)$ compared to the long-term strategy's base line, not containing the effect of z_t. The experimental results with respect to Thai stocks and US stocks are shown in Table 4.3 and Table 4.4, respectively.

Table 4.3 Experimental results in profit percentage of our Bayesian trading strategies comparing to the benchmark buy-and-hold strategy on Stock Exchange of Thailand (SET). Bold face numbers indicate the cases where our method is superior. In the last row, "W/D/L" summarizes Win/Draw/Lose of our method compared to the benchmark.

Symbol	Long-term Thresholds				Medium-term Thresholds				Buy and Hold
	5%	10%	15%	20%	3%	5%	7%	10%	
CPALL	**58.40%**	38.74%	**49.74%**	**56.42%**	26.92%	42.59%	32.08%	36.34%	44.46%
CPN	30.67%	30.67%	30.67%	30.67%	-1.06%	-6.39%	-4.32%	14.04%	30.67%
EASTW	9.16%	6.18%	4.29%	4.29%	**10.49%**	3.93%	**15.07%**	**20.64%**	9.16%
GLOW	-11.78%	-11.78%	-11.78%	-11.78%	**-3.24%**	-14.16%	**-10.66%**	**-3.13%**	-11.78%
HMPRO	31.88%	**47.93%**	**58.00%**	0%	**51.69%**	**69.87%**	**86.49%**	8.87%	32.28%
QH	-24.08%	-24.08%	-24.08%	**-9.70%**	**5.73%**	**0.49%**	**4.08%**	**2.31%**	-24.08%
ROBINS	36.28%	36.28%	36.28%	36.28%	32.70%	**36.77%**	**45.64%**	**36.97%**	36.28%
SCB	-17.46%	-17.46%	**-14.10%**	**-9.08%**	**-11.67%**	**-14.09%**	**-14.74%**	**4.79%**	-17.46%
SNC	-4.45%	-4.45%	-7.00%	**-0.91%**	**18.46%**	**8.20%**	**8.02%**	**8.20%**	-4.45%
TTW	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**-1.31%**	**-0.50%**	**3.67%**	**0%**	-3.73%
Average	**10.86%**	**10.20%**	**12.19%**	**9.62%**	**12.92%**	**12.67%**	**16.55%**	**12.90%**	9.14%
W/D/L	2/7/1	2/6/2	4/4/2	5/3/2	7/0/3	6/0/4	8/0/2	7/0/3	
	13/20/7				28/0/12				

Table 4.4 Experimental results in profit percentage of our Bayesian trading strategies comparing to the benchmark buy-and-hold strategy on US stock markets: NYSE and NASDAQ.

Symbol	Long-term Thresholds				Medium-term Thresholds				Buy and Hold
	5%	10%	15%	20%	3%	5%	7%	10%	
WMT	**6.07%**	**13.00%**	**23.04%**	**0.00%**	**20.47%**	**5.97%**	**12.02%**	**0.00%**	-11.98%
HD	10.56%	28.71%	28.71%	28.71%	**48.35%**	**40.77%**	27.55%	27.48%	28.71%
KO	6.06%	6.06%	6.06%	6.06%	**20.61%**	**17.24%**	0.00%	0.00%	6.06%
G	5.33%	26.50%	26.50%	26.50%	**37.40%**	**29.74%**	21.22%	1.14%	26.50%
AAPL	5.19%	5.19%	5.19%	5.19%	**18.69%**	**22.83%**	**41.27%**	**26.84%**	6.77%
NKE	12.72%	12.72%	12.72%	12.72%	**22.53%**	**25.00%**	**24.50%**	**32.96%**	12.72%
BK	**26.36%**	**23.72%**	1.42%	1.42%	**5.52%**	**6.44%**	-3.68%	**4.84%**	1.42%
CF	**27.16%**	**15.87%**	-53.74%	-53.74%	**-17.60%**	**-16.84%**	**-36.21%**	**-41.05%**	-53.74%
CSCO	12.00%	**23.00%**	20.48%	20.48%	8.42%	10.98%	14.98%	20.00%	20.48%
DIS	0.82%	0.82%	0.82%	0.82%	**14.37%**	**12.17%**	-1.45%	-8.47%	0.82%
Average	**11.23%**	**15.56%**	**7.12%**	**4.82%**	**17.88%**	**15.43%**	**10.02%**	**6.37%**	3.78%
W/D/L	3/3/4	4/5/1	1/8/1	1/8/1	9/0/1	9/0/1	4/0/6	5/0/5	
	9/24/7				27/0/13				

From Tables 4.3 and 4.4, we can see that in the SET market, the average profit percentages are 10.86%, 10.20%, 12.19%, 9.62%, 12.92%, 12.67%, 16.55%, 12.90% for different thresholds in long-term strategy and medium-term strategy which are more than 9.14% from the buy-and-hold strategy. In the US market, the average profit percentages are 11.23%, 15.56%, 7.12%, 4.82%, 17.88%, 15.43%, 10.02%, 6.37% for different thresholds in long-term strategy and medium-term strategy which are more than 3.78% from the buy-and-hold strategy.

In the total of 80 trading simulations on SET firms, our method results in greater performance 41 times, while the results based on buy-and-hold strategy in better performance 19 times (the remaining 20 times are draws). Similarly, in the total of 80 trading simulations on NYSE and NASDAQ firms, our method results in greater performance 36 times, while the results based on buy-and-hold strategy in better comparison 20 times (the remaining 24 times are draws). Summing up results of markets in the two countries, our method outperforms the benchmark 77 times, yet underperforms only 39 times. These are promising results where we shall analyze statistically significance of the results more formally in the next subsection. Here, we shall firstly interpret and discuss the experimental results in Tables 4.3 and 4.4 in detail.

From those two tables, it can be seen that there are 44 draws, which occur only in the cases of the long-term trading strategy. All 44 draws happen because of the exact same reason: our model predicts undervaluation at the beginning of the testing period. i.e. the first observed PE falls deeply below the base line PE^* (exceeding the specified threshold). After that, the observed PE is never once able to overly exceed PE^* with respect to the given threshold, i.e. $PE^*(1 + Tr)$ is quite high in the test set so that no selling is possible. Therefore, in this case, our trading behaves exactly just like the benchmark buy-and-hold strategy. Note that there is no draw in the results of the medium-term trading strategy. The reason is because the estimated medium-term noisy effect z_t makes the base line $PE^*(1+z_t)$ move near to the observed PE which results in more frequent trading.

Disregarding the draws, our long-term trading strategy still beats the benchmark with 22 wins versus 14 loses. This is mainly due to the volatility of the observed PE in most stocks so that our strategy of buying in an undervalued price and selling in an overvalued price with respect to PE^* is possible. However, it is not the case that trading induced from our model constantly outperforms the benchmark. For the case of the so-called growth stocks (Lynch and Rothchild, 2000), i.e. stocks with consistently increasing earnings and price, it is not so easy for our model to beat the benchmark. If the threshold is set too low, our method will lead to buying and selling early, and thus results in less profit. See Figure 4.5 for example. If the threshold is set too high, our method may lead to buying when the price is already high or lead to doing nothing at all because it is never undervalued with respect to the specified threshold. Another special case where our method fails to beat the benchmark is when there are the so-called non-recurring earnings, i.e. extra incomes which occur only once and should not be taken into account in the calculation of fundamental value. In this case, the market knows that these extra earnings are temporary and do not give it credit, i.e. the price does not go up according to this profit. Our method has not taken this information into account and thus is fooled to believe that a stock is undervalued. See Figure 4.6.

On the other hand, the results of our method equipped with medium-term trading strategy show impressive superiority, 55 wins versus 25 loses to the benchmark. The key factor of success is its tracking ability of the medium-term noisy effect z_t by our filtering algorithm presented in Section 3.3. When the new base

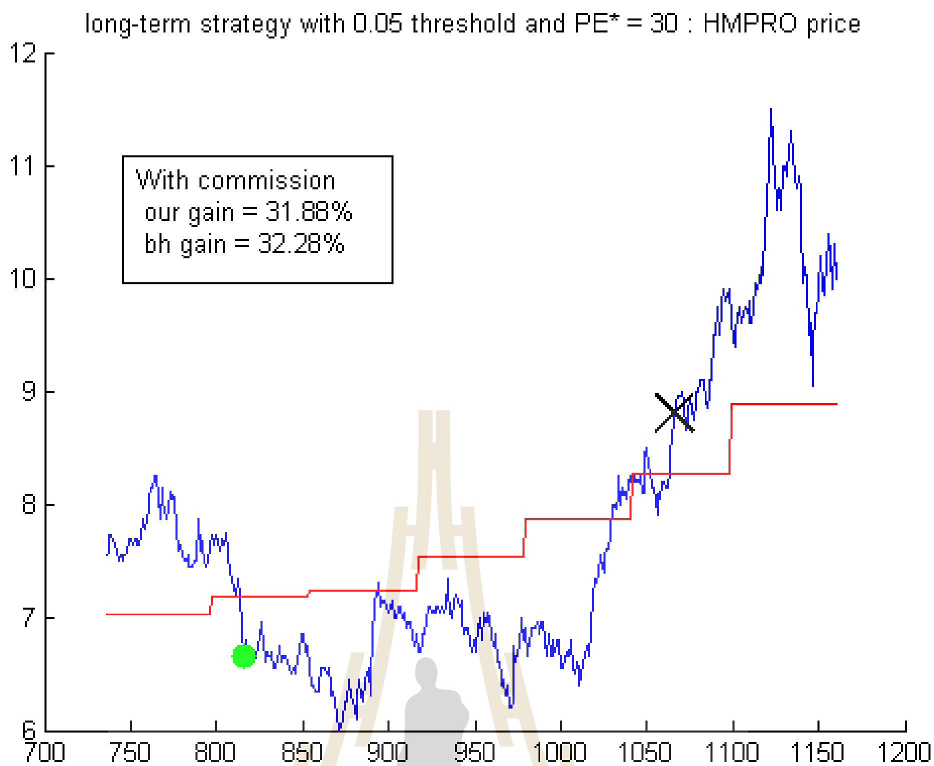

Figure 4.5 Example of a a Growth stock (Lynch and Rothchild, 2000) where its earnings are consistently increasing. For a growth stock, if the threshold size is non-optimal, we may buy and sell too early (and do not make trading frequently enough) which results in less profit.

line $PE^*(1+z_t)$ is predicted accurately, undervalued and overvalued prices are also accurately detected and so the probability of our profitable trading is increasing. Nevertheless, with more frequent trading, commission fees increase substantially and sometimes can significantly reduce our performance as shown in Figure 4.2.

Figure 4.7 and Figure 4.8 show an example of long-term trading strategy and medium-term trading strategy of WMT, respectively. In Figure 4.7, The red line denotes the profit of buy-and-hold strategy, while the blue line denotes the profit of the long-term strategy based on our Bayesian inference results. We

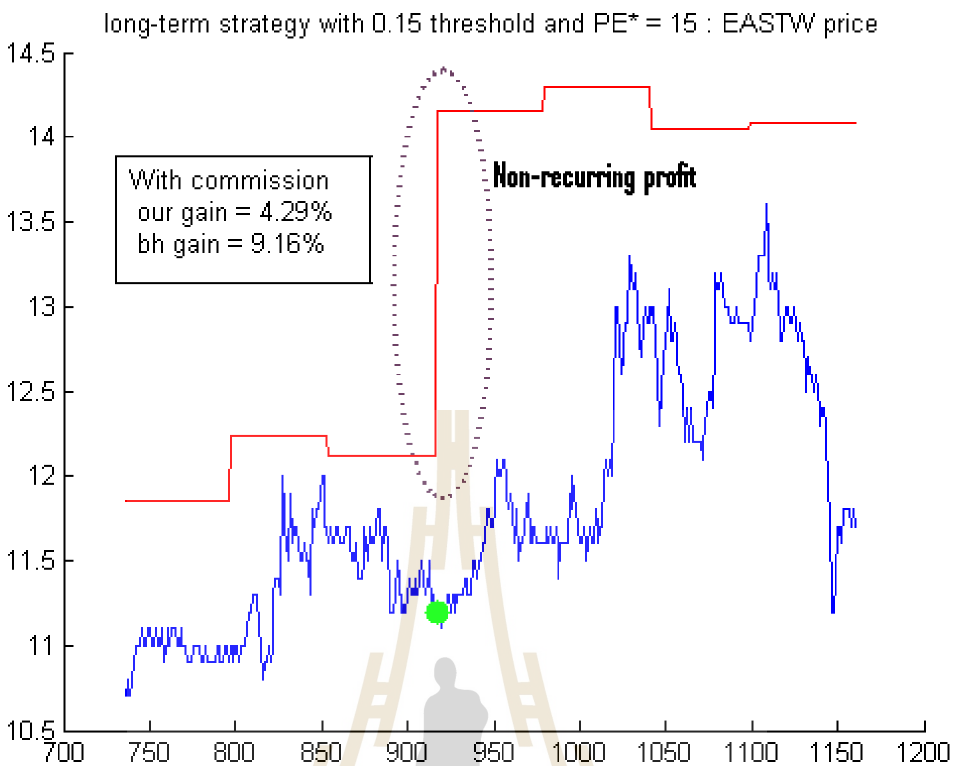

Figure 4.6 Example of Non-recurring earnings and Non-recurring profit. Since our fundamental price calculation is naive and does not take into account the fact that these earnings are temporary, it estimates a too-high fundamental price due to this extra earnings.

use the stock of WMT as an example with 4 different thresholds 0.05, 0.1, 0.15 and 0.2. In Figure 4.8, the red line denotes the profit of buy-and-hold strategy, while the blue line denotes the profit of the medium-term strategy based on our Bayesian inference results. We use the stock of WMT as an example with 4 different thresholds 0.03, 0.05, 0.07 and 0.1. From these two figures, we can see that our strategy can beat buy-and-hold strategy in most of the 2-year testing set.

For the stock WMT, in the 2-year testing set, the average profit percentage of the long-term strategy are 6.35%, 8.99%, 13.92%, 0% for different thresholds, respectively. The average profit percentage of the medium-term strategy are 5.5%,

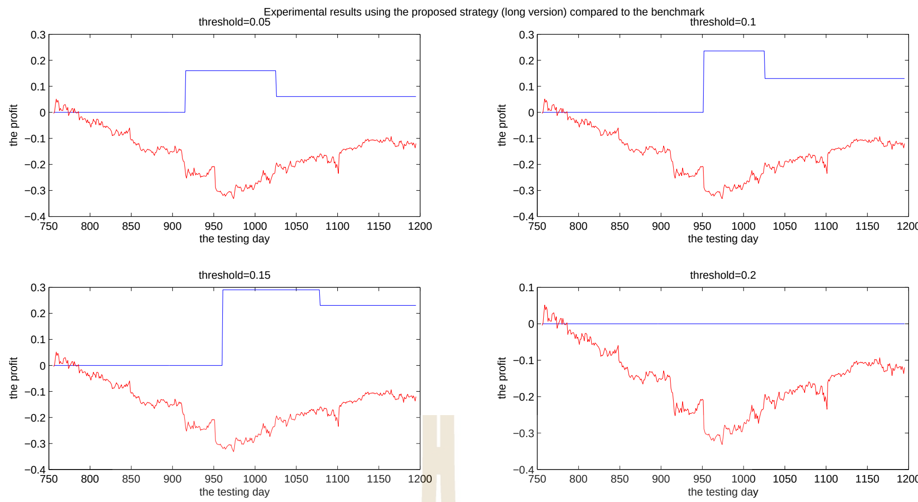

Figure 4.7 Experimental results using the long-term strategy comparing to the benchmark buy-and-hold strategy.

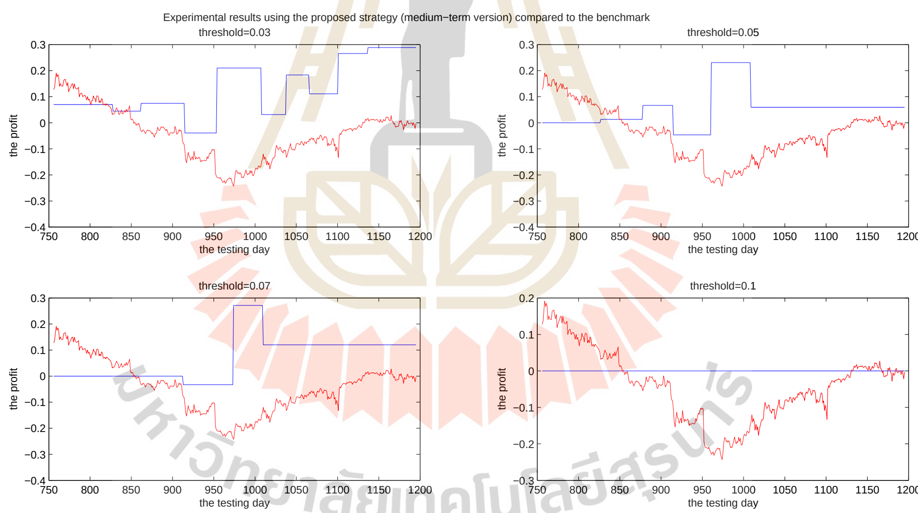

Figure 4.8 Experimental results using the medium-term strategy comparing to the benchmark buy-and-hold strategy.

5.24%, 6.9%, 0% for different thresholds, respectively. The average profit percentage of the buy-and-hold strategy is -15.41%. This shows that our trading strategy for WMT consistently outperforms the buy-and-hold strategy on average, see Figure 4.9. In Figure 4.9, "bh" represents "the buy-and-hold strategy",

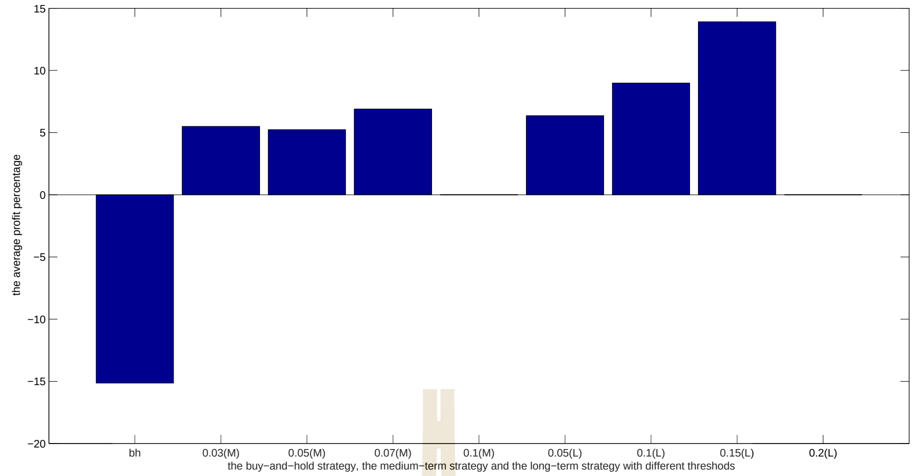

Figure 4.9 The average profit percentage of WMT, using the buy-and-hold strategy, the medium-term strategy and the long-term strategy with different thresholds.

"M" represents "the medium-term strategy" and "L" represents "the long-term strategy".

4.3.2 Portfolio level experiments

In this subsection, to more realistically simulate a real-world individual investor, we construct a portfolio of stocks and test its performance against the benchmark. Here, we use a rule-of-thumb commonly employed in practice saying that a good portfolio should consists of around 15 stocks (which contradicts the mainstream theory (D. Domian, 2007)). For individual value investors who believe they can beat the market by analyzing each firm carefully, usually, they do not feel comfortable to hold too many stocks (like 100 stocks recommended by academic financial literature) because investors need time to update and analyze the information of all their stock holdings.

To test the performance of a 15-stock portfolio of our method against the benchmark, we employ the method of bootstrap resampling (Horowitz, 1997). For each boostrap sample, a set of 15 stocks are selected randomly from Tables 4.3 and 4.4 to form an equally-weighted portfolio. We are interested in the difference in performance between our method and the benchmark on each boostrap sample. After all bootstrap samples are drawn, we can also estimate the average difference in performance between the two methods. More precisely, let X be a random variable representing difference in % profit between our model and the benchmark (our % profit minus the benchmark's % profit). By repeating the bootstrap resampling 10,000 times, we are able to construct an empirical distribution of X. This empirical distribution allows us to calculate $E[X]$, the average % profit difference between the two methods, and $Pr(X \geq 0)$, the probability that our method has superior or equal performance to the benchmark. In addition to a portfolio consisting of stocks from the two markets, we also test portfolio performance from SET or US alone. Since the number of stocks considered in this work shown in Tables 4.3 and 4.4 is 10, a 7-stock portfolio is constructed for these cases instead of a 15-stock portfolio. The results are shown in Table 4.5.

From Table 4.5, our method beats the benchmark on every case on average (since $E[X] > 0$ for all cases). However, on a single-country portfolio, about half of the cases have the confidence levels of superiority $Pr(X \geq 0)$ less than 80%. Most of them have satisfactory confidence levels greater than 70% though. On the other hand, on a two-country 15-stock portfolio, although $E[X]$ is roughly an average of the two single-country portfolios, $Pr(X \geq 0)$ is significantly increasing so that most cases have the confidence levels of superiority greater than 80%, half of them is greater than 90%. This statistically confirms the superiority of our method over the benchmark on selected stocks. This phenomenon of confidence-level increasing is due to the diversification effect on portfolio with a higher number of stocks. Finally, we note that the empirical distribution of X is usually skewed and long-tail as illustrated in Figure 4.10. In this non-simple probability distribution case, bootstrap empirical-distribution estimation employed in the present thesis usually provides more accurate result than traditional analytical asymptotic estimations (Horowitz, 1997).

Table 4.5 Experimental results in portfolio level testing. X denotes a random variable representing difference in % profit between our model and the benchmark. The distribution of X is estimated using the method of Bootstrap Resampling. Bold face denotes a case where there is more than 80% confidence that our method is more superior or equal to the benchmark.

Portfolio	Long-term Thresholds				Medium-term Thresholds			
	5%	10%	15%	20%	3%	5%	7%	10%
SET								
$E[X]$	**1.70%**	1.07%	**3.12%**	0.45%	3.73%	3.52%	**7.46%**	3.70%
$Pr(X \geq 0)$	**86.76%**	66.72%	**84.99%**	57.45%	71.16%	69.53%	**81.44%**	74.28%
US								
$E[X]$	7.30%	**11.88%**	3.35%	1.03%	**14.04%**	**11.75%**	**6.22%**	2.62%
$Pr(X \geq 0)$	72.25%	**97.06%**	72.51%	72.29%	**99.82%**	**99.78%**	**88.48%**	70.35%
SET+US								
$E[X]$	**4.62%**	**6.42%**	**3.19%**	0.77%	**8.89%**	**7.55%**	**6.87%**	**3.18%**
$Pr(X \geq 0)$	**80.64%**	**97.47%**	**92.78%**	66.04%	**97.94%**	**96.30%**	**92.74%**	**80.58%**

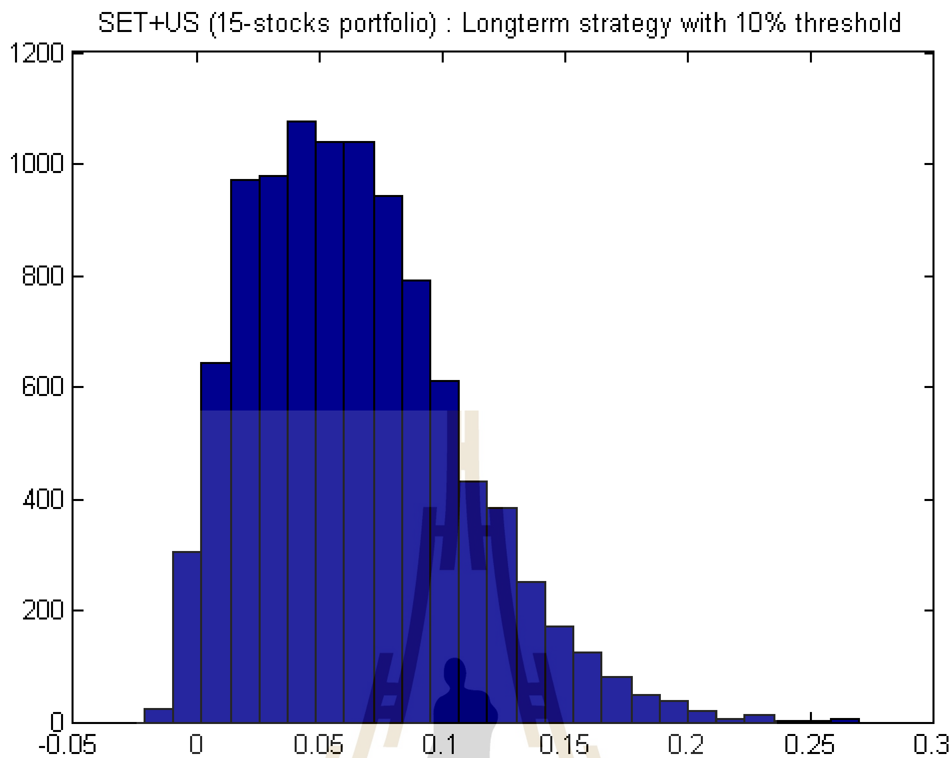

Figure 4.10 A histogram illustrated common empirical distributions of X with right-skewed and long-tail. In this specifical illustration, our method is equipped with the long-term trading strategy with 10% threshold.

CHAPTER V

STOCK TRADING BASED ON THE GORDON GROWTH MODEL

We formally reformulate the stock price dynamics model to reflect an important variable which is equity risk premium. Equity risk premium is an important input in corporate finance and valuation. We present stock trading using PE ratio in Chapter III and Chapter IV. Some researchers use the earnings to price ratio related to equity risk premium, see (Zarowin, 1990; Ball, 1992; Jaffe, Keim, and Westerfield, 1989). However, it is not a good proxy for equity risk premium. In this chapter, we focus on estimating equity risk premium on firm-level stocks based on Gordon Growth model as explained in Section 2.1.

5.1 Background of fundamental investment based on the Gordon Growth model

As explained in Section 2.1, the Gordon Growth model is described as follows,

$$P_t^* = \frac{D_t(1+g)}{r_f + \pi - g}, \tag{5.1}$$

whereP_t^* denotes the value of stock at the end of period t, D_t is the yearly dividend in the tth year, $t = 1, 2, ...,$ r_f is risk free rate, g is a constant growth rate and π is equity risk premium which is normally assumed to be a constant in a period.

For a risk free security, there are two kinds of security, one is short-term government security which is called treasury bill. Another is long-term government security which is called treasury bonds. Some practitioners and a surprising number of academics and textbooks have a logic that there is no risk in the treasury bill and they use the treasury bill rate as the risk free rate. Whereas, in corporate finance and valuation, people use the long-term government bond rate as the risk free rate not the short-term rate.

The equity risk premium reflects risk in an economy. When the equity risk premium rises, investors charge a higher price for risk and will pay lower prices for the stock. There are three broad approaches to estimating the equity risk premium.

One is the survey approach. In this approach, a group of investors and managers are investigated to get the equity risk premium. The challenge of this approach is to find a subset of investors and managers who best represent the stock market. For different subsets, the equity risk premium seems to be different. (Kaustia, etc. 2011) surveyed 1465 investors and the estimated equity risk premiums from male advisors are lower than those from female advisors. With the development of stock market and technology, investors can do various surveys and then give their estimations of equity risk premiums. However, these estimations for equity risk premiums are not best reflections of good forecasts in the future. (Bartholdy and Peare, 2005) present that the performance of Capital Asset Pricing Model (CAPM) and three-factor model (3FM) is poor: they explain on average 0.03, 0.05 of differences in returns, respectively. They are even in the wrong direction. (Fisher and Statman, 2000) prove the negative relationship between investor sentiment and stock returns.

Nevertheless, it is important that these estimations from experienced investors and managers can be considered partly or completely in our DBN model.

The second is historical approach which is the most widely used approach.

In this approach, we estimate historical returns on stocks over long periods. The equity risk premium is the difference between the historical return on stocks and actual return on bonds (usually government security). To estimate historical returns on stocks, there are two averaging approaches which are the arithmetic average and the geometric average. The arithmetic average return is the mean of the series of annual returns, whereas the geometric average return is calculated by Eq.(5.2).

$$geometricaverage = (\frac{value_T}{value_1})^{1/T} - 1 \tag{5.2}$$

where $value_1$ is the value at the start of the period, $value_T$ is the value at the end of the period. (Indro and Lee 1997) compare these two historical returns, and find them both wanting. Many estimation services and academics argue that the arithmetic average is the best estimate of the equity risk premium. However, in corporate finance and valuation, more people use geometric average as the estimate of the equity risk premium.

The consensus for estimating for future risk premiums is that historical approach is the best way (Damodaran, 2016). However, since we need enough long historical data in this approach, the approach is not feasible in emerging markets. Using historical approach, if we use a shorter and more recent time period on firm-level stocks to get more updated estimate, the cost is the great standard error for the equity risk premium. If we use the entire data on firm-level stocks, the equity risk premiums may have little relevance to today's firm. In fact, even with some modifications, the historical approach is a backward looking premium.

The third is the implied approach, which is used to estimate a forward-looking equity risk premium. We can illustrate the implied equity risk premium with the dividend discount model (DDM) as explained in Section 2.1. The Gordon

Growth model is simple version of DDM. (Claus and Thomas, 2001) argue that the expected return estimates from fundamentals (the implied approach) are more precise than returns which are from the survey approach. (Fama and French, 2002) support this conclusion with their new results. The goal is to support that fundamentalists determine the equity risk premium by Bayesian analysis. We also combine historical data with expert beliefs, such as the beliefs using the survey approach. In the next section, we will formalize the stock price dynamics to estimate the equity risk premium for the future based on Gordon Growth model.

5.2 Model the stock price dynamics

Due to behavioral volatility, the stock price deviates from its value. We formalize the process and divide the temporary effects into two categories as explained in Chapter III. Based on Gordon Growth model and the phenomenaon called mean reversion, we have

$$P_t = \frac{D_t(1+g)}{r_f + \pi - g}(1+z_t)(1+\varepsilon_t) \tag{5.3}$$

where

(a) P_t is the stock price and D_t is dividend at the end of period t;

(b) r_f is risk free rate which is constant;

(c) π is the variable for equity risk premium which is assumed to be constant;

(d) g is a dividend growth rate which is constant;

(e) z_t is the variable for medium-term effect, which persist for a period and we model z_t as a Markov chain;

(f) ε_t is the variable for short-term effect and $\varepsilon_t \sim N(0, \sigma^2)$.

Mathematically,

$$\ln(P_t/D_t) = \ln \frac{(1+g)}{r_f + \pi - g}(1+z_t) + \ln(1+\varepsilon_t) \tag{5.4}$$

Let $y_t = \ln(P_t/D_t)$. Since ε_t is usually small, it can be approximated by $\ln(1+\varepsilon_t) \approx \varepsilon_t$. Then we then have the model

$$y_t = \ln \frac{(1+g)}{r_f + \pi - g}(1+z_t) + \varepsilon_t \tag{5.5}$$

Note that the historical data for stock price and dividend are available from stock market and thus y_t is observable. π and z_t are hidden variables. π is constant and z_t is time-varying. We assume that π and z_t are discrete. We represent the model using Bayesian network, see Figure 5.1.

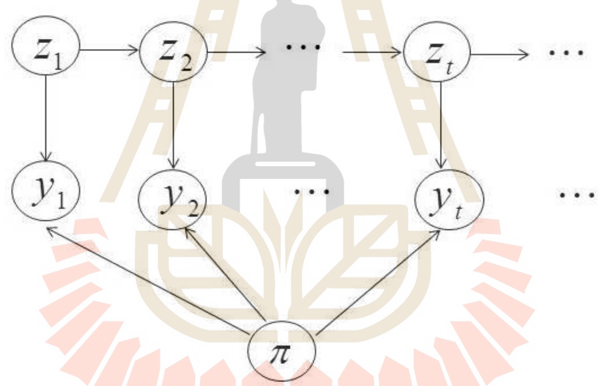

Figure 5.1 The proposed model represented by the DBN. y_t is an observable quantity, while π and $\{z_t\}$ are unobservable.

The structure of the model based on Gordon Growth model is the same to the structure based on PE ratio as explained in Chapter III. To derive mathematical equations for inference and parameter estimation in our model, we also first define the probability distribution functions as in Section 3.2, which are the transition probability distribution function, the emission probability distribution function and the initial probability distribution function. For the transition probability distribution function (pdf), it is the same as in Section 3.2. Next, we define the emission probability distribution function by Eq.(5.6) and the initial probability distribution function by Eq.(5.7) and Eq.(5.8) as follows.

For all $m \in \{1, ..., M\}$, $n \in \{1, ..., N\}$,$t \in \{1, 2, ...\}$

$$p(y_t|z_t = a_m, \pi = b_n) \triangleq \phi_{mn}(y_t). \tag{5.6}$$

By Eq.(5.4), $\phi_{mn}(y_t) = N(\ln(\dfrac{(1+g)(1+a_m)}{r_f + \pi - g}), \sigma^2)$. The matrix $\Phi_t = (\phi_{mn})_{M\times N}$ is the emission matrix at period t.

For each $m \in \{1, ..., M\}$,

$$u_m \triangleq p(z_1 = a_m), \tag{5.7}$$

where $0 \leq u_m \leq 1$ and $\sum_{m=1}^{M} u_m = 1$.

For each $n \in \{1, ..., N\}$,

$$v_n \triangleq p(\pi = b_n), \tag{5.8}$$

where $0 \leq v_n \leq 1$ and $\sum_{n=1}^{N} v_n = 1$. The vectors $\mathbf{u} = (u_m)_M$ and $\mathbf{v} = (v_n)_N$ are the initial vectors.

Therefore, the set of model parameters is $\theta = \{\mathbf{W},\mathbf{u},\mathbf{v}, \sigma^2\}$ which is similar to the set in Section 3.2. Derivation of Bayesian inference and parameter estimation based on GGM are similar to the derivation based on PE ratio as explained in Section 3.2.

5.3 Experiment based on the Gordon Growth model

In this section, we make stock trading simulations in US (NYSE and NASDAQ) market which represent matured stock markets. We select firms from US market which own at least 5 years historical trading data including stock price and stock dividend. We collect daily 5-year (Mar 13, 2012 to Sep 30, 2016) historical data for each firm consisting of 1147 data for stocks.

The historical price and dividend are denoted by $P_1, ..., P_{1147}$ and $D_1, ..., D_{1147}$, respectively. Stock prices are adjusted for stock splitting and stock dividending. The yearly dividend is defined by the summation of the latest four quarterly dividends. The first 3-year historical data (Mar 13, 2012 to Dec 31,2014) $P_1, ..., P_{706}$ and $D_1, ..., D_{706}$ are used as a training data to learn parameters using EM algorithm and get $z_1, ..., z_{706}$ by the method of smoothing. Here, we use government bond as our risk free rate. The variables g, the possible values $a_1, ..., a_M$ of z_t and the possible values $b_1, ..., b_N$ of π are set by experts.

Each trading simulation is conducted for each individual stock with the remaining 2-year historical data (Jan 1, 2015 to Sep 30, 2016) $P_{707}, ..., P_{1147}$ and $D_{707}, ..., D_{1147}$ to measure the performance of both our model and the benchmark. There are two versions which are long-term strategy and medium-term strategy. Details for strategies can be found in Section 4.1. The main idea for the trading strategy is similar to the strategy in Section 4.1. The medium-term strategy is emphasized here. The reason is that the results of our method based on PE ratio equipped with medium-term strategy show impressive superiority. The strategy is described in the following.

Let I_t and N_t be available cash and total shares at date t, respectively. Initially, $I_{706} = I$ and $N_{706} = 0$. The trading strategy can be defined now. For each date t, exactly one of the following cases holds:

(i) $P_t/D_t \leq A_t(1 - Tr)$ and $I_t > 0$ (buy-low case) where $Tr \in (0, 1)$ is a threshold, $A_t = \frac{1+g}{r_f+\pi-g}$ for the long-term strategy and $A_t = \frac{1+g}{r_f+\pi-g}(1 + z_t)$ for the medium-term strategy. In this case, buy the stock with all cash, so that $N_{t+1} = C.I_t/P_T$ and $I_{t+1} = 0$.

(ii) $P_t/D_t \geq A_t(1 + Tr)$ and $I_t = 0$ (sell-high case). In this case, sell all the holding stock to get cash $I_{t+1} = P_t.N_t.C$ and $N_{t+1} = 0$.

(iii) If case (i) and case (ii) are not satisfied, do nothing. So, $I_{t+1} = I_t$ and $N_{t+1} = N_t$. At the end of a trading simulation $t = 1147$, the total profit is $I_{1147} + P_{1147}.N_{1147} - I_{706}$, so that we can compare with the buy-and-hold strategy profit.

We test 6 (more than 4) different thresholds for medium-term trading strategy. The experimental results with respect to US stocks using medium-term strategy are shown in Table 5.1.

Table 5.1 Experimental results in profit percentage of the medium-term strategy comparing to the benchmark buy-and-hold strategy on US stocks.

Symbol	Medium-term Thresholds						Buy and Hold
	3%	5%	7%	9%	12%	15%	
WMT	**-6.9%**	**0.00%**	**17.80%**	**10.45%**	**11.82%**	**14.79%**	-11.98%
NKE	**16.72%**	**21.42%**	3.66%	8.26%	12.72%	12.72%	12.72%
BK	**31.05%**	**42.92%**	**66.69%**	**71.99%**	1.42%	1.42%	1.42%
CSCO	6.60%	13.69%	16.69%	**21.54%**	4.55%	20.48%	20.48%
Average	**11.87%**	**19.51%**	**26.21%**	**28.06%**	**7.63%**	**12.35%**	5.66%
W/D/L	3/0/1	3/0/1	2/0/2	3/0/1	1/2/1	1/3/0	
	13/5/6						

From Table 5.1, we can see that in the total of 24 trading simulations on US stocks, our methods results in greater performance 13 times, while buy-and-hold strategy results in better 6 times. The remaining 5 times are draws. It can be seen that these 5 draws occur only in the cases that the threshold is 12% or 15%. These two thresholds are bigger than other thresholds. The threshold which is not over 10% is better for medium-term strategy. Disregarding the draws, our medium-term trading strategy beats the benchmark with 13 wins versus 6 loses.

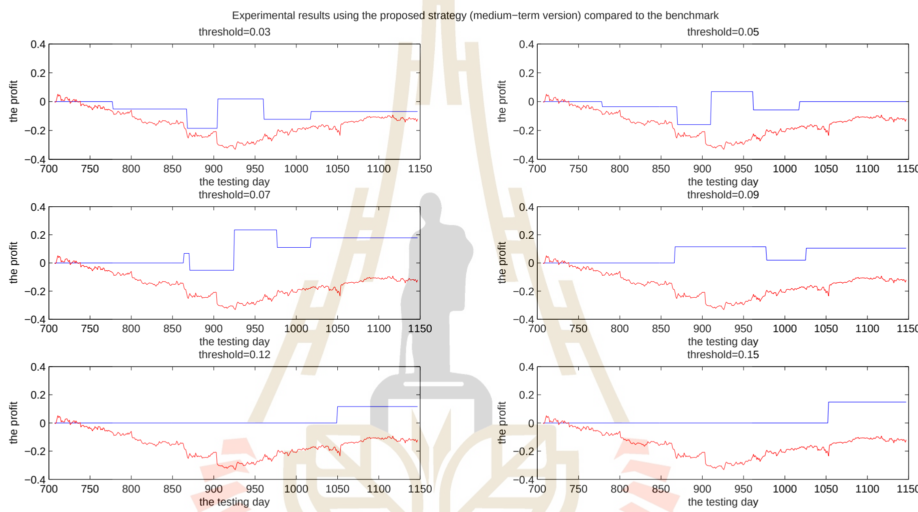

Figure 5.2 Experimental results using the medium-term strategy comparing to the benchmark buy-and-hold strategy.

We also use WMT as an example, see Figure 5.2. In Figure 5.2, the red line denotes the profit of buy-and-hold strategy, while the blue line denotes the profit of the proposed strategy based on our Bayesian inference results. We use the stock of WMT as an example with 6 different thresholds 0.03, 0.05, 0.07, 0.09, 0.12, 0.15. We can see that we can beat the buy-and-hold strategy. Here we use medium-term trading strategy with 6 different thresholds. From the figure, we can see that the propose strategy (medium-term version) can beat buy-and-hold strategy.

CHAPTER VI

CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this thesis, we propose to apply the advanced Dynamic Bayesian Network (DBN) methodology on firm-level stocks. Fundamentalists have clear financial interpretations. In the stock market, there is a phenomenon called mean reversion. Based on this phenomenon and fundamental value, we model stock price dynamics based on PE ratio and the Gordon Growth model, respectively. In the framework, we also combine the expert information with the historical data. We use Bayesian network to simplify the calculation of our model and we have derived both Bayesian inference (forward-backward algorithm) and parameter estimation (EM-MAP) algorithms.

Based on the results of our model, a simple but practical strategy is invented. Next, we do experiments based on PE ratio and Gordon Growth model, respectively. We make stock trading simulations based on PE ratio on Thailand stocks and US stocks. We collect 5 years daily data for 20 firms from various industries. We consider two versions of our strategy: long-term strategy and medium-term strategy. For each trading strategy, we choose 4 different thresholds. For those two versions of trading strategy, our method always beats the gold buy-and-hold strategy. The results with medium-term strategy show more superiority. Experiments in both individual firm-level and portfolio level show statistically significant superiority of our method.

Also, we make stock trading simulations based on Gordon Growth model on US stocks. Based on the results using PE ratio, we only use medium-term trading strategy to do experiments. For medium-term trading strategy, we choose 6 different thresholds and our method also beats the gold buy-and-hold strategy.

The study can be used as a decision support system to investment experts, or used to construct a trading strategy directly as illustrated in Chapter IV. This model is most suitable for one majority category of practitioners, namely, value investors in a security market (our model is not suitable for technical investors and mainstream academic investors). However, there exists a limitation in our model. If the earnings are negative, our model cannot work.

6.2 Future work

There are many possible future directions for the present work. The first direction is to more formally reformulate the stock price dynamics model to reflect other important economic and financial variables, e.g. interest rate, return of equity. It is possible to make our DBN model more realistic by allowing time-varying short-term noisy effect, the so-called dynamic volatility (Wu et al., 2013) or allowing dynamic volume (Llorente et al., 2002). Another promising direction in behavioral finance which can be taken into account in our model is the topic of heterogeneous agents (Hommes, 2013). To improve our inference procedure, approximate inference such as Variational Bayes (Murphy, 2012), or stochastic inference such as Markov chain Monte Carlo (Bishop, 2006) are very promising future directions.

We use experiments to show that our trading strategy based on our model outperforms the buy-and-hold strategy. The core of the trading strategy is to sell the overvalued security and buy the undervalued one. We have proposed the strategy based on calculating the fundamental value of a security. Whereas pairs trading strategy is to use relative pricing to determine that a security is overvalued or undervalued. Thus, those two strategies are well compared.

REFERENCES

REFERENCES

Agarwal, V., and Naik, N. Y. (2004). Risks and portfolio decisions involving hedge funds. **Review of Financial Studies**. 17(1): 63–98.

Ball, R. (1992). The earnings-price anomaly. **Journal of Accounting and Economics**. 15(2-3): 319–345.

Bartholdy, J., and Peare, P. (2005). Estimation of expected return: CAPM vs. Fama and French. **International Review of Financial Analysis**. 14(4): 407–427.

Bishop, C. M. (2006). **Pattern Recognition and Machine Learning**. New York: Springer Science Press.

Black, F. (1986). Noise. **The Journal of Finance**. 41(3): 528–543.

Bloomfield, R., Libby, R., and Nelson, M. (2000). Underreactions, overreactions and moderated confidence. **Journal of Financial Markets**. 3(2): 113–137.

Bodie, Z. (2009). **Investments**. Tata McGraw-Hill Education.

Bondt, W. F., and Thaler, R. (1985). Does the stock market overreact? **The Journal of Finance**. 40(3): 793–805.

Brealey, R. A., Myers, S. C., Allen, F., and Mohanty, P. (2012). **Principles of Corporate Finance**. Tata McGraw-Hill Education.

Campbell, J. Y., and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. **Review of Financial Studies**. 1(3): 195–228.

Carhart, M. M. (1997). On persistence in mutual fund performance. **The Journal of Finance**. 52(1): 57–82.

Carvell, S. A., Sullivan, T., and Pari, R. (1989). **Analysts Forecast and**

Price/Earning Ratios. Cornell University: School of Hotel Administration Collection.

Chen, A.-S., Leung, M. T., and Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the taiwan stock index. **Computers and Operations Research**. 30(6): 901–923.

Claus, J., and Thomas, J. (2001). Equity premia as low as three percent? evidence from analysts' earnings forecasts for domestic and international stock markets. **The Journal of Finance**. 56(5): 1629–1666.

Cooper, M. J., Dimitrov, O., and Rau, P. R. (2001). A rose.com by any other name. **The Journal of Finance**. 56(6): 2371–2388.

Cowen, A., Groysberg, B., and Healy, P. (2006). Which types of analysts firms are more optimistic? **Journal of Accounting and Economics**. 41(1–2): 119–146.

Damodaran, A. (2012). **Investment Valuation,** $3^{rd}Ed.$. LifeVest Publishing, Inc.

Damodaran, A. (2016). **Equity Risk Premiums (ERP): Determinants, Estimation and Implications–The 2016 Edition.**

D. Domian, M. R., D. Louton. (2007). Diversification in portfolios of individual stocks: 100 stocks are not enough. **The Financial Review**. 42(4): 557–570.

Dechow, P. M., Hutton, A., and Sloan, R. (2000). The relation between analysts' forecasts of long-term earnings growth and stock price performance following equity offering. **Journal of Contemporary Accounting Research**. 17(1): 1–32.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. **Journal of Political Economy**. 98(4):

703–738.

Ellison, A. M. (2004). Bayesian inference in ecology. **Ecology Letters**. 7(6): 509–520.

Enke, D., and Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. **Expert Systems with Applications**. 29(4): 927–940.

Fama, E. F. (1976). **Foundations of Finance**. Hwa-Tai.

Fama, E. F. (1991). Efficient capital markets. **The Journal of Finance**. 46(5): 1575–1617.

Fama, E. F., and French, K. R. (2002). The equity premium. **The Journal of Finance**. 57(2): 637–659.

Fisher, K. L., and Statman, M. (2000). Investor sentiment and stock returns. **Financial Analysts Journal**. 56: 16–23.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). **Bayesian Data Analysis**. New York: CRC Press.

Hassan, R. (2009). A combination of hidden markov model and fuzzy model for stock market forecasting. **Neurocomputing**. 72(16): 3439–3446.

Heaton, J., and Lucas, D. (1999). Stock prices and fundamentals. **NBER Macroeconomics Annual**. 14: 213–242.

Henry, E., Robinson, T., Stowe, J., and Pinto, J. (2010). **Equity Asset Valuation, 2^{nd} ed.** CFA Institute.

Hommes, C. (2013). **Behavioral Rationality and Heterogenous Expectation in Complex Economic Systems**. Cambridge University Press.

Horowitz, J. L. (1997). Bootstrap methods in econometrics: theory and numerical performance. **Econometric Society Monographs**. 28: 188–222.

Jaffe, J., Keim, D. B., and Westerfield, R. (1989). Earnings yields, market values, and stock returns. **The Journal of Finance**. 44(1): 135–148.

Kao, L., Chiu, C., and Lu, C. (2013). A hybrid approach by integrating wavelet-based feature extraction with mars and svr for stock index forecasting. **Decision Support Systems**. 54(3): 1228–1244.

Kazem, A., Shariff, E., and Hussain, F. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting. **Applied Soft Computing**. 13(2): 947–958.

Khakzad, N., Khan, F., and Amyotte, P. (2011). Safety analysis in process facilities: Comparison of fault tree and bayesian network approaches. **Reliability Engineering and System Safety**. 96(8): 925–932.

Kim, M. J., Nelson, C. R., and Startz, R. (1991). Mean reversion in stock prices? a reappraisal of the empirical evidence. **The Review of Economic Studies**. 58(3): 515–528.

Lu, C., Lee, T., and Chiu, C. (2009). Financial time series forecasting using independent component analysis and support vector regression. **Decision Support Systems**. 47(2): 115–125.

Lynch, P., and Rothchild, J. (2000). **One Up on Wall Street, 2^{nd} Ed.** Simon and Schuster.

Malkiel, B. (2016). **A Random Walk Down Wall Street: The Time-tested Strategy for Successful Investing,** $11^{th}Ed.$. W.W. Norton and Company.

Manzan, S. (2007). Nonlinear mean reversion in stock prices. **Quantitative and Qualitative Analysis in Social Sciences**. 1(3): 1–20.

Mark, H. (2011). **The Most Important Things: Uncommon Sense for the Thoughtful Investor,** $2^{nd}Ed.$. Columbia University Press.

Miller, M. H., Muthuswamy, J., and Whaley, R. E. (1994). Mean reversion of standard and poor's 500 index basis changes: Arbitrage-induced or statistical illusion? **The Journal of Finance**. 49(2): 479–513.

Murphy, J. (1999). **Technical Analysis of The Financial Markets: A Comprehensive Guide to Trading Methods and Applications**. Penguin.

Murphy, K. P. (2012). **Machine Learning: A Probabilistic Approach**. MIT Press.

Poterba, J. M., and Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. **Journal of Financial Economics**. 22(1): 27–59.

Pound, J., and Zeckhauser, R. (1990). Clearly heard on the street: The effect of takeover rumors on stock prices. **The Journal of Business**. 63(3): 291–308.

Preis, T., Moat, H. S., and Stanley, H. E. (2013). Quantifying trading behavior in financial markets using google trends. **Scientific Reports**. 3: 1–6.

Rashes, M. (2001). Massively confused investors making conspicuously ignorant choices. **The Journal of Finance**. 56(5): 1911–1927.

S. Boyd, L. V., S. J. Kim, and Hassibi, A. (2007). A tutorial on geometric programming. **Optimization and Engineering**. 8(1): 67–127.

Shannon, B. (2008). **Technical Analysis Using Multiple Time Frames**. LifeVest Publishing, Inc.

Shleifer, A., and Vishny, R. (1997). The limits of arbitrage. **Journal of Finance**. 52(1): 35–55.

Strelioff, C. C., Crutchfield, J. P., and Hübler, A. W. (2007). Inferring Markov chains: Bayesian estimation, model comparison, entropy rate, and out-of-class modeling. **Physical Review E**. 76(1): 011106.

Szyszka, A. (2013). **Behavioral Finance and Capital Markets**. New York:

Palgrave Mackmillan.

Tumarkin, R., and Whitelaw, R. F. (2001). News or noise? internet postings and stock prices. **Financial Analysts Journal**. 57(3): 41–51.

Valstar, M. F., and Pantic, M. (2007). Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. In **International Workshop on Human-Computer Interaction** (pp. 118–127).

Wen, Q., Yang, Z., Song, Y., and Jia, P. (2010). Automatic stock decision support system based on box theory and svm algorithm. **Expert Systems with Applications**. 37(2): 1015–1022.

Yeh, C., Huang, C., and Lee, S. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. **Expert System and Applications**. 38(3): 2177–2186.

Zarowin, P. (1990). What determines earnings-price ratios: Revisited. **Journal of Accounting, Auditing and Finance**. 5(3): 439–454.

APPENDIX

APPENDIX

```
% Note that this program is about the historical data. p=load('cpall 5.txt');
p=p(1160:-1:1);
p(1:80)=p(1:80)/2;
ee=load('cpall earnings.txt');
ee=ee(22:-1:1);
ee(1:5)=ee(1:5)/2;
for i=1:19
eee(i)=ee(i)+ee(i+1)+ee(i+2)+ee(i+3);
end
nn=[62 58 63 62 62 59 63 61 62 59 62 62 61 57 63 62 63 57 62];
T=sum(nn);
for i=1:19
nnn(i)=sum(nn(1:i));
end
e(1:nnn(1))=eee(1);
for i=2:19
e(nnn(i-1)+1:nnn(i))=eee(i);
end
e=e';
for t=1:T
pe(t)=p(t)/e(t);
y(t)=log(p(t)/e(t));
end
```

```
p=y';
a=[-0.2 -0.1 0.1 0.2];
M=4;
b=[27:44];
N=18;
% Note that this program is the function which is to use EM algorithm to
estimate the parameter u.
function yu=emu(sigma,y,M,N,a,b,u,v,W,T)
for m=1:M
for n=1:N
Mean(m,n)=log(b(n)*(1+a(m)));
py(m,n)=normpdf(y(1),Mean(m,n),sigma);
al(m,n)=py(m,n)*u(m)*v(n);
end
end
ppy1=py;
AA1=al;
c(1)=sum(sum(al));
alpha1=al./c(1);
for t=2:T
for m=1:M
for n=1:N
py(m,n)=normpdf(y(t),Mean(m,n),sigma);
end
end
ppyt=py;
```

```
A=py.*(W*alphat-1);
AAt=A;
c(t)=sum(sum(A));
alphat=A./c(t);
end
betaT-1=W'*ppyT./c(T);
for t=T-2:-1:1
betat=W'*(ppyt+1.*betat+1)./c(t+1);
end
for t=1:T-1
altat=alphat.*betat;
end
yu=sum(alta1');
% Note that this program is the function which is to use EM algorithm to
estimate the parameter v.
function yv=emv(sigma,y,M,N,a,b,u,v,W,T)
for m=1:M
for n=1:N
Mean(m,n)=log(b(n)*(1+a(m)));
py(m,n)=normpdf(y(1),Mean(m,n),sigma);
al(m,n)=py(m,n).*u(m)*v(n);
end
end
ppy1=py;
c(1)=sum(sum(al));
alpha1=al./c(1);
```

```
for t=2:T
for m=1:M
for n=1:N
Mean(m,n)=log(b(n)*(1+a(m)));
py(m,n)=normpdf(y(t),Mean(m,n),sigma);
end
end
ppyt=py;
A=py.*(W*alphat-1);
c(t)=sum(sum(A));
alphat=A./c(t);
end
betaT-1=W'*ppyT./c(T);
for t=T-2:-1:1
betat=W'*(ppyt+1.*betat+1)./c(t+1);
end
for t=1:T-1
altat=alphat.*betat;
end
k=zeros(1,N)+1;
yv=(sum(alta1)+(k-1))./(1+sum(k-1));
% Note that this program is the function which is to use EM algorithm to
estimate the parameter W.
function yW=emW(sigma,y,M,N,a,b,u,v,W,T)
for m=1:M
for n=1:N
```

```
Mean(m,n)=log(b(n)*(1+a(m)));
py(m,n)=normpdf(y(1),Mean(m,n),sigma);
al(m,n)=py(m,n).*u(m)*v(n);
end
end
ppy1=py;
c(1)=sum(sum(al));
alpha1=al./c(1);
for t=2:T
for m=1:M
for n=1:N
Mean(m,n)=log(b(n)*(1+a(m)));
py(m,n)=normpdf(y(t),Mean(m,n),sigma);
end
end
ppyt=py;
A=py.*(W*alphat-1);
c(t)=sum(sum(A));
alphat=A./c(t);
end
betaT-1=W'*ppyT./c(T);
for t=T-2:-1:1
betat=W'*(ppyt+1.*betat+1)./c(t+1);
end
D1=zeros(M,M);
for t=2:T-1
```

```
Dt=Dt-1+betat.*ppyt*alphat-1'./c(t);
end
DT=DT-1+ppyT*alphaT-1'./c(T);
W=W.*DT;
for m=1:M
for mm=1:M
if m==mm
PW(m,mm)=0.997;
else
PW(m,mm)=0.001;
end
end
end;
W=(W+1000*PW)./(repmat(sum(W),[M,1])+1000*ones(M,M));
yW=W;
```

CURRICULUM VITAE

NAME : Haizhen Wang

DATE OF BIRTH : August 26, 1985

NATIONALITY : Chinese

EDUCATION BACKGROUND :

- B.Sc.(Mathematics), Tianjin University of Technology and Education, China, July 2007.

- M.Sc.(Applied Mathematics), Tianjin University of Technology and Education, China, March 2010.

PUBLICATIONS :

- Haizhen Wang, Ratthachat Chatpatanasiri and Pairote Sattayatham (2017). Stock trading using PE Ratio: A dynamic Bayesian Network modeling on behavioral finance and fundamental investment.Thai Journal of Mathematics. (Under Review)