Faculty of Engineering and InformationTechnology
University of Technology, Sydney

Cross-market Behavior Modeling

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Wei Cao

June 2015

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

To My Parents and Xin

for your love and support

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Longbing Cao for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I also would like to appreciate my co-supervisor Associate Prof. Jian Zhang for providing me with continuous support throughout my PhD study and research. Without his professional guidance and persistent help, this thesis would not have been possible.

I thank my fellow labmates in Advanced Analystics Institute: Yin Song, Junfu Yin, Xuhui Fan for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three years.

I place on record, my sense of gratitude to the team leader, Zhigang Zheng in Australian Taxation Office for his expert and sincere help in the project. I am also grateful to my team members, Mu Li, Chunming Liu, David Wei and Allen Lin, for their hard working in project teams.

Last but not the least, I would like to thank my family: my parents and my boyfriend, for their unconditional support, both financially and emotionally throughout the whole PhD studying.

Wei Cao

December 2014 @ UTS

Contents

List of Figures

List of Tables

List of Publications

Papers Published

- **Wei Cao**, Liang Hu, Longbing Cao (2015). Deep Modeling Complex Couplings within Financial Markets. *in* 'Proceedings of the the Twenty-Ninth AAAI Conference on Artificial Intelligence (**AAAI-15**)', full paper accepted.

- Zhigang Zheng, Wei Wei, Chunming Liu, **Wei Cao**, Longbing Cao, and Maninder Bhatia (2015). An Effective Contrast Sequential Pattern Mining Approach on Taxpayer Behavior Analysis, **World Wide Web (2015)**, pp.1-19.

- **Wei Cao**, Longbing Cao (2014). Financial Crisis Forecasting via Coupled Market State Analysis. **Intelligent Systems, IEEE**, 30 (2), pp.18-25 .

- Liang Hu, **Wei Cao**, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu (2014). Bayesian Heteroskedastic Choice Modeling on Non-identically Distributed Linkages. *in* 'Proceedings of the IEEE International Conference on Data Mining (**ICDM 2014**)', short paper accepted.

- Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, **Wei Cao** (2014). Deep Modeling of Group Preferences for Group-Based Recommendation. *in* 'Proceedings of the the Twenty-Eight AAAI Conference on Artificial Intelligence (**AAAI-14**)', full paper accepted.

- **Wei Cao**, Longbing Cao, Yin Song (2013). Coupled Market Behavior Based Financial Crisis Detection. The 2013 International Joint Conference on Neural Networks (**IJCNN2013**)' , pp. 1-8.

- **Wei Cao**, Cheng Wang, Longbing Cao (2012). Trading Strategy Based Portfolio Selection for Actionable Trading Agent. Agents and Data Mining Interaction (**ADMI 2012**), pp. 191-202.

- Yin Song, Longbing Cao, Xuhui Fan, **Wei Cao**, Jian Zhang. Characterizing A Database of Sequential Behaviors with Latent Dirichlet Hidden Markov Models. arXiv:1305.5734v1 [stat.ML].

Papers to be Submitted/Under Review

- **Wei Cao**, Longbing Cao (2015). Coupled Cross-Market Behavior Analysis for Forecasting Financial Market Trends, to be submitted as a journal paper.

- **Wei Cao**, Longbing Cao. Multi-layer Coupled Hidden Markov Model for Cross-market Behavior Analysis and Trend Forecasting, to be submitted as a journal paper.

Research Reports of Industry Projects

- **Wei Cao**, Zhigang Zheng, Wei Wei, Maninder Bhatia. Action Response Model of Activity Statement – data mining modelling and evaluation report, Debt Collection and Optimisation Project, Australian Taxation Office, Nov 2013.

- **Wei Cao**, Zhigang Zheng and Wei Wei. Due-date Self-finalising Model of Income Tax – data mining modelling and evaluation report, Debt Collection and Optimisation Project, Australian Taxation Office, May 2014.

Abstract

During the 2007 global financial crisis which was triggered by subprime borrowers in the US mortgage markets, strong market linkages were observed between different financial markets. The sharp fluctuations in the global stock market, commodity market and interest market illustrate some of the coupled behaviors that exist between various markets, namely the crisis effect is passed from one market to another through couplings. Here coupled behaviors refer to the activities (e.g. changes of market indexes) of financial markets which are associated with each other in terms of particular relationships. Therefore, a good understanding of coupled behaviors is of great importance in cross-market applications such as crisis detection and market trend forecasting. For instance, if the coupled behaviors are properly understood and modeled, investors can predict financial crisis and avoid the big loss, by detecting the changes of coupled relations between financial crisis period and non-crisis period.

However, understanding and modeling coupled behaviors is quite challenging for following reasons: (1) The various coupled structures across financial markets (e.g. coupled relations between different types of markets, and coupled relations between the same type of market in different countries) bring challenges in terms of understanding and modeling them. (2) Various types of couplings. The typical forms are intra-coupling, inter-coupling and temporal-coupling. (3) The complex interactions between markets are driven by hidden features which cannot be observed directly from observation/data. (4) Different applications in cross-market analysis lead to the consideration

of input factors/variables selection.

All of these challenges the existing methods for cross-market analysis, which can be roughly categorized into two types: time series analysis represented by Logistic regression, Autoregressive Integrated Moving Average (ARIMA) and Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) models. Model-based methods explore machine models such as Artificial Neural Networks (ANN) and Hidden Markov Models (HMM). The main limitations lie in their deficiencies: (1) Existing approaches are usually focused on the simple correlations of the cross-market, rather than coupled behaviors between markets. (2) State-of-the-art research work is usually built directly from the observation/data. Hidden features behind the observation/data are often ignored or only weakly addressed. (3) Some approaches follow assumptions that are too strong to match real financial markets.

Based on the above research limitations and challenges, this thesis reports state-of-the-art advances and our research innovations in understanding and modeling complex coupled behaviors for the purpose of cross-market analysis.

Chapter 3 presents a new approach, called Coupled Market Behavior Analysis (CMBA) for financial crisis detection. This caters for nonlinear couplings between major indicators selected from different markets, and it detects different coupled market behaviors at crisis and non-crisis periods. Chapter 4 seeks to overcome the limitations of most current methods which conduct financial crisis forecasting directly through observation and overlook the hidden interactions between markets. In this chapter, Coupled Market State Analysis (CMSA) is presented to build forecasters based on coupled market states instead of observation.

Chapter 5 reports a new approach for market trend forecasting by analyzing its hidden coupling relationships with different types of related financial markets. Chapter 6 proposes Hierarchical Cross-market Behavior Analysis (HCBA) to forecast a stock market's movements, by exploring the complex coupling relationships between variables of markets from a country (Layer-1 coupling) and couplings between markets from various countries (Layer-2

coupling). In addition, Chapter 7 designs a Coupled Temporal Deep Belief Network (CTDBN) which accommodates three different types of couplings across financial markets: interactions between homogeneous markets from various countries (intra-market coupling), interactions between heterogeneous markets (inter-market coupling) and interactions between current and past market behaviors (temporal coupling). With a deep-architecture model to capture the high-level coupled features, the proposed approach can infer market trends.

In terms of cross-market applications (i.e. financial crisis detection and market trend forecasting), our proposed approaches and frameworks for modeling coupled behaviors across financial markets outperform state-of-the-art methods from both technical and business perspectives. All of these outcomes provide insightful knowledge for investors who naturally seek to make profits and avoid losses. Accordingly, cross-market behavior modeling is a promising research topic with lots of potential for further exploration and development.

Chapter 1

Introduction

This thesis explores complex coupled behaviors across different financial markets. What distinguishes this thesis from previous work is not only understanding various coupled structures across financial markets and building coupled models to analyze the different kinds of coupled behaviors, but also, applying coupled behavior analysis into cross-market applications including financial crisis detection and market trends forecasting. This chapter describes the background information of coupled behaviors in financial markets, challenges and limitations, research issues and contributions of our work.

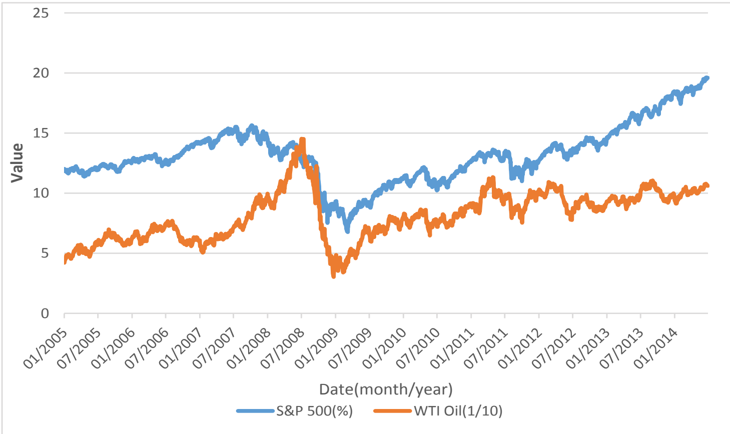
1.1 Background

Since the development of economic globalization, global financial markets became more integrated with the flow of capital across the world. As a result, new information arising in one market affects not only its future behaviors (for example, disappointing news can lead to stock market downtrends), but also the future behaviors of other markets (Chan, Treepongkaruna, Brooks & Gray 2011). We call these transmissions cross-market behaviors, namely there exist coupled behaviors across different financial markets.

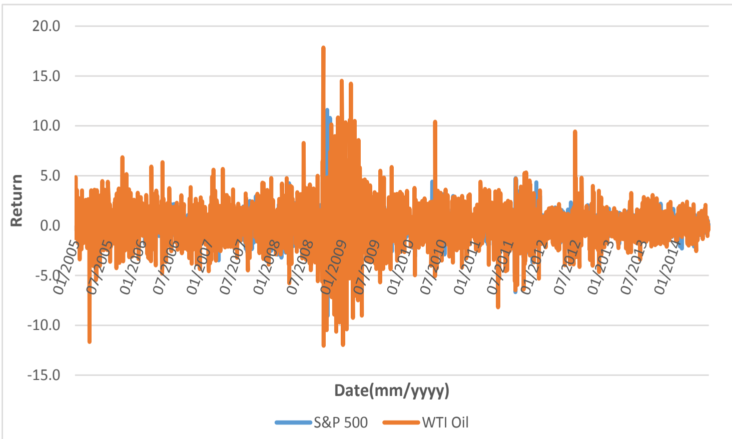
The coupled behaviors between various markets can be detected more clearly in financial crisis periods, where the breakdown in one market spills over to the rest of the world. We all know that global financial markets suffered catastrophic losses from 2007 to 2009. This was primarily triggered by the bailout of Bear Stearns' hedge funds in 2007 and the bankruptcy of Lehman Brothers in September 2008 (Longstaff 2010). Strong market linkages can be found through the changes of different market indexes. For example, in the most volatile year 2008, the S&P 500 index (connect to the stock market) dropped from 1267.38 in July to 896.24 in November, while the WTI Oil price (connect to the commodity market) reached a record high of 147 dollars per barrel in July and dropped to 60 dollars in November.

These examples show some of the coupled behaviors of financial markets in crisis periods. The relations can be illustrated more clearly in Figure 1.1. This displays data from January 2005 to June 2014 in two typical markets: the US stock market and commodity market. From Figure 1.1 (a) we find that the relationships between the two markets are very complex, and they fluctuated sharply during the financial crisis period (e.g. 2007-2009). Figure 1.1 (b) shows the relationships between the returns of the two markets, where the return is calculated by $R_t = \frac{P_t - P_{t-1}}{P_{t-1}} * 100\%$, and R_t and P_t are, respectively, the return and closing price at time t. It can clearly be seen that the returns of the two markets reveal similar changes, which means that they are highly correlated. In addition, the two return series show higher fluctuations during the crisis period as compared to other normal periods.

The above analysis supports the existence of coupled behaviors across financial markets. Coupled market relations should be paid attention to from both the trading and risk management point of view. For example, if the coupled behaviors can be properly defined and modeled, investors may profit from correctly predict the market trends, or detect impending financial crisis for the purposes of avoiding losses. All these factors motivate research on modeling complex coupled behaviors in cross-markets which constitutes the topic of this thesis.



(a) Indexes of Stock Market and Commodity Market



(b) Returns of Stock Market and Commodity Market

Figure 1.1: Demonstrations of Index Series

1.2 Challenges and Limitations

1.2.1 Research Challenges

As stated above, financial markets are strongly linked to each other. Therefore, understanding and modeling the complex coupled behaviors across markets is far from easy. The main challenges are as follows:

- **Various coupled structures across financial markets.** There are plenty of financial markets in the real world, and the markets are coupled for various reasons. There are three typical coupled structures: a) same type of markets in different countries (e.g. the US stock market and Chinese stock market), this is a relatively simple structure and has been investigated by many researchers (Baker & Hart 2008). b) different types of markets (e.g. the stock market and commodity market) c) mixed markets, namely different types of markets in various countries.

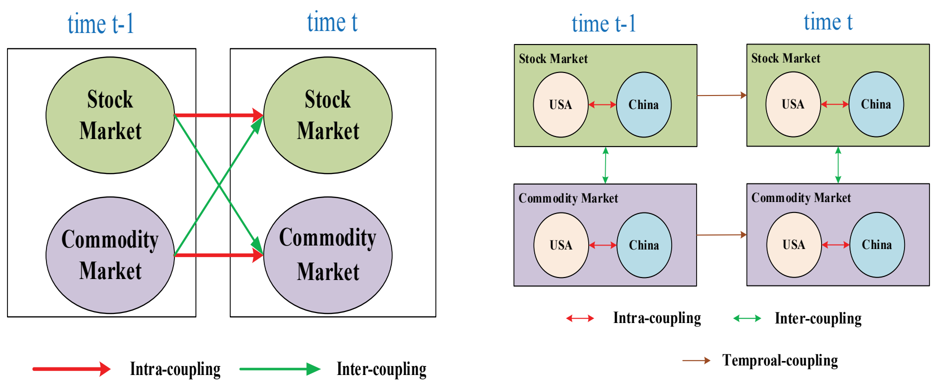
  Modeling and analysis should not limited to simple correlations between the same type of markets (Chen, Firth & Meng Rui 2002, Yang, Min & Li 2003), but also to following two more complex cases: a) the complex coupled behaviors between different types of markets; and b) the interactions between mixed markets. Figure 1.2 proposes examples of the two structures of coupled markets. Figure 1.2 (a) illustrates an example of coupled behaviors between two different types of markets: the stock market and commodity market, Figure 1.2 (b) proposes an example of interactions between two different markets (the stock market and commodity market) in two countries (the USA and China), and they are all coupled with each other. In conclusion, the different structures of coupled markets bring challenges to understanding and modeling them.

- **Various types of couplings.** As illustrated in Figure 1.2, in order to capture the complex coupled behaviors, we need to explore the following three types of couplings across financial markets:

  a) intra-coupling. It refers to the interactions between the same or similar markets. For example, the intra-coupling in Figure 1.2 (a) is represented by correlations from the same stock market, while in Figure 1.2 (b) it is illustrated by interactions from same type of markets: the Chinese stock market and USA stock market.

b) inter-coupling. It refers to the interactions with distinctions. For example, as illustrated in Figure 1.2 (a), the interactions between the stock market and commodity market are represented by inter-coupling.

c) temporal-coupling. It describes the transitional influences from past behaviors.



(a) A Demonstration of Coupled Behaviors between Different Types of Markets

(b) A Demonstration of Coupled Behaviors between Mixed Markets

Figure 1.2: Examples of Different Coupled Structures across Markets

- **Hidden characteristics.** This denotes that the complex interactions between market behaviors are driven by hidden features which cannot be observed directly from observation/data such as market indexes (Chan et al. 2011). Once the drivers (i.e. hidden features) are changed, the observations fluctuate. Accordingly, an understanding of the hidden features (which may be abstract) should be taken into consideration.

- **Application issue.** The target of understanding and modeling coupled behaviors in cross-market is to help investors obtain more profits and avoid loss. However, in different applications the selection of input factors/variables should also be paid attention to. For example, if one is engaged in financial crisis detection, it is important to select different macroeconomic indicators (e.g. indexes from global commodity market and interest market) rather than indexes in various countries, since the changes in the interactions between macroeconomic indicators are more significant (Chevallier 2012, Candelon, Piplack & Straetmans 2009).

In order to find the gap between the challenges and existing cross-market analysis approaches, we examine the current most widely used approaches and their limitations in following section.

1.2.2 Current Research Limitations

In previous studies, many different approaches have been proposed to learn the cross-market.

Many theoretical methods have been proposed to verify the existence of cross-market correlations/linkages, such as Unit Root Tests (Phillips & Perron 1988), Vector AutoRegression (Hamori & Imamura 2000, Aktan, Kopurlu et al. 2009), CoIntegration Tests (Johansen & Juselius 1990) and Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) (Bodart & Reding 1999). However, these studies only provide the evidence of cross-market correlations. In the real world it is much more important to work out real financial problems such as financial crisis detection and market trend forecasting by using these correlations.

The approaches applied in real cross-market problems (e.g. market trend forecasting) can be roughly categorized into two groups: a) time series analysis which has been widely applied and explored in financial markets. The typical models include Logistic regression (Laitinen & Laitinen 2001), Autoregressive Integrated Moving Average (ARIMA) (Contreras, Espinola, Nogales & Conejo 2003) and GARCH (Marcucci 2005) models. b) model-based approaches which explore machine models such as Artificial Neural Networks (ANN) (Pan, Tilakaratne & Yearwood 2005) and Hidden Markov Models (HMM) (Hassan & Nath 2005).

Despite great progress and development, limited efforts have been made in deep understanding and modeling of the complex coupled behaviors across financial markets. Below, we summarize and list the main limitations and challenges of current research into cross-market analysis while detailed introductions and evaluations of the related work are given in Chapter 2.

- Existing approaches usually focus on the simple correlations of the cross-market, rather than coupled behaviors between markets. As illustrated above, there are different types of couplings between financial markets, namely intra-coupling, inter-coupling and temporal coupling. Such complex couplings challenge the existing methods which simply analyze the independent behavior from market observation/data. However, a deep exploration of coupled behaviors is necessary for us to understand how market behaviors are coupled with each other, and to find how the coupled behaviors act under different market situations. For example, the changes of couplings between financial crisis and non-crisis periods can support the correct detection of an impending financial crisis.

- State-of-the-art research work is usually built directly from observation/data. Hidden features under the observation/data are often ignored or only weakly addressed. However, we need to understand what such hidden factors are since they are the drivers of the complex interactions.

- Some approaches follow assumptions that are too strong to match the real financial markets. Financial markets, especially the couplings, are very complex and reveal non-linear characteristics, which means that it is difficult for some approaches with linear assumption (e.g. the Logistic approach). In addition, the market time series are non-stationary, especially during the crisis period, and this may challenge some approaches with stationary assumption (e.g. the ARIMA approach).

This thesis aims to break through the limitations and address the challenges listed above. It also seeks to introduce new approaches and novel frameworks to understand, model and learn from cross-market behaviors.

1.3 Research Issues

Based on the aforementioned research challenges and limitations, the key research issues associated with this thesis are summarized in Table 1.1, and they are discussed in terms of following three aspects:

- **Understanding different coupled structures across financial markets.** As illustrated above, there are various coupled structures. For a specific case, in order to learn from the complex hidden coupled relations, we first need to focus on developing behavior-oriented specifications and formalization to describe the specific coupled structures (e.g. the coupled relations between different types of markets). This provides a unified and formalized mechanism for describing and presenting behavior interactions, desired requirements or properties, and behavior impacts and patterns.

- **Modeling and learning different types of couplings with hidden features.** With the formalization of coupled behaviors under a specific coupled structure, the research addresses the task of couplings (i.e. intra-, inter- and temporal coupling) modeling and learning. The modeling process includes building proper machine learning models which can capture the different types of complex hidden couplings, and then mapping the formalized behavior analysis into specific models. The models can be developed from the Coupled Hidden Markov Model (CHMM), State Space Model (SSM) and Deep Belief Networks (DBN) which are introduced in detail in Chapter 2. The proposed models are then used to learn the hidden couplings and infer future market behaviors by corresponding inference and parameter learning processes.

Table 1.1: Research Issues in Each Chapter

Research Issues	Detailed Research Issues	Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Understanding Various Coupled Structures	Different Types of Markets	✓	✓	✓		
	Different Types of Markets in Various Countries				✓	✓
Modeling and Learning Couplings	Intra-coupling	✓	✓	✓	✓	✓
	Inter-coupling	✓	✓	✓	✓	✓
	Temporal coupling					✓
Evaluating Coupled Behavior Analysis	Financial Crisis Detection	✓	✓			
	Market Trend Forecasting			✓	✓	✓

- **Evaluating coupled behaviors analysis across markets.** The quantitative research targets coupled behavior evaluation by applying in crisis detection and market trend forecasting in the real cross-market, through exposing the coupled interactions between different types of financial markets, or between different markets in various countries. The performance can then be evaluated from the technique perspective (e.g. accuracy and precision) and the business perspective (e.g. rate of return) to exhibit whether the coupled behavior analysis across financial markets is helpful and effective or not.

1.4 Research Contributions

This thesis mainly focuses on resolving two typical real cross-market problems: crisis detection and market trend forecasting, by modeling the complex hidden couplings across financial markets with different coupled structures. We also build two novel frameworks for crisis detection, and three approaches for market trend forecasting. The main contributions are listed as follows:

- The proposal of a new approach, called *Coupled Market Behavior Analysis*, to detect financial crisis by catering for the often nonlinear couplings between major indicators selected from different global financial markets (Chapter 3);

- The exploration of a *Coupled Hidden Markov Model* to characterize the coupled market behaviors of stock, commodity and interest markets as case studies (Chapter 3);

- The proposal of a framework called *Coupled Market State Analysis* to predict financial crisis. This approach predicts crisis based on the coupled market states over different types of financial markets, and it avoids the vulnerability of data in traditional observation-based approaches (Chapter 4);

- The creation of a *Coupled State Space Model* over coupled market data to learn the coupled market states behind the data (Chapter 4);
- The development of a new forecasting algorithm to infer one market trends by forecasting its probability distributions, through analyzing its hidden couplings with different types of other related financial markets (Chapter 5);
- The proposal of a framework of *Hierarchical Cross-market Behavior Analysis* to predict market trends, by exploring the complex couplings from two layers: Layer-1 coupling represents the relations between variables of markets from a country and Layer-2 coupling denotes the interactions between markets from various countries (Chapter 6);
- The creation of a *Multi-layer Coupled Hidden Markov Model* to capture the complex hierarchical couplings between various markets in different countries (Chapter 6);
- The proposal of a *Coupled Temporal Deep Belief Network* to build a hierarchical architecture of observations, from which we can deep model the coupled hidden features between different financial markets in various countries. This avoids the vulnerability of observation in traditional approaches (Chapter 7);
- The exploration of a *Conditional Restricted Boltzmann Machine* at the bottom level to learn the intra-market coupling among homogeneous market and the design of a *Conditional Gaussian Restricted Boltzmann Machines* at the top level to disentangle high-level inter-market coupling from heterogeneous markets based on intra-market coupling (Chapter 7).

1.5 Thesis Structure

The thesis is structured as follows:

Chapter 2 provides the literature review of the definition of behavior, coupled behavior and coupled behavior analysis. The foundation of methods used in this thesis are introduced, which include the coupled hidden markov model, state space model, deep belief network. In addition, we review two typical cross-market analysis problems: financial crisis detection and market trend forecasting.

Chapter 3 introduces a new financial crisis detection approach to consider the nonlinear characteristics in financial markets and the couplings between major indicators selected from different markets, called coupled market behavior analysis. A CHMM-based model is then built to capture the different coupled behaviors between the financial markets during crisis and non-crisis periods. The approach is deployed to detect the 2007 subprime mortgage crisis by selecting major indicators from the commodity, stock and interest markets. The results show the clear advantage of coupled market behavior analysis against the Signal, Logistic and ANN based methods with a significant accuracy improvement.

Chapter 4 proposes coupled market state behavior analysis to predict financial crisis, by modeling the hidden interactions between different types of financial markets. In particular, a coupled state space model is built to capture the coupled hidden market states across markets. The experiments with real financial data show the clear advantage of the coupled market state behavior analysis based approach against the state-of-the-art observation based methods.

Chapter 5 develops a new forecasting approach to achieve improved performance for forecasting financial market movements by analyzing the complex coupling relationships between various indicators in different financial markets. Through capturing the coupled market behaviors, a coupled hidden markov model is built to infer a market trend by forecasting its probability distributions. Experimental results on nine years of real financial market data show that the proposed approach outperforms other baselines, from both technical and business perspectives.

Chapter 6 introduces a new market trend forecasting approach by analyzing the multi-layered, hidden couplings between various markets in different countries. In the chapter, a hierarchical coupled cross-market behavior analysis framework and a MCHMM are proposed to capture the complex hierarchical coupling relationships between various markets in different countries. The approach is then applied to infer market trends with 10 years of data from two markets (the stock market and currency market) in 13 countries as case studies.

Chapter 7 proposes a deep learning approach to capture the three types of underlying complex couplings across multiple financial markets: intra-market coupling represents interactions between homogeneous markets from various countries, inter-market coupling represents the interactions between heterogeneous markets and temporal coupling illustrates the interactions between current and past market behaviors. A coupled deep belief network is then designed to accommodate the above three types of couplings across financial market behaviors. With the deep-architecture model to capture the high-level coupled features, the proposed approach can infer market trends.

Chapter 8 concludes the thesis and outlines the scope for future work.

Figure 1.3 shows the research profile of this thesis.

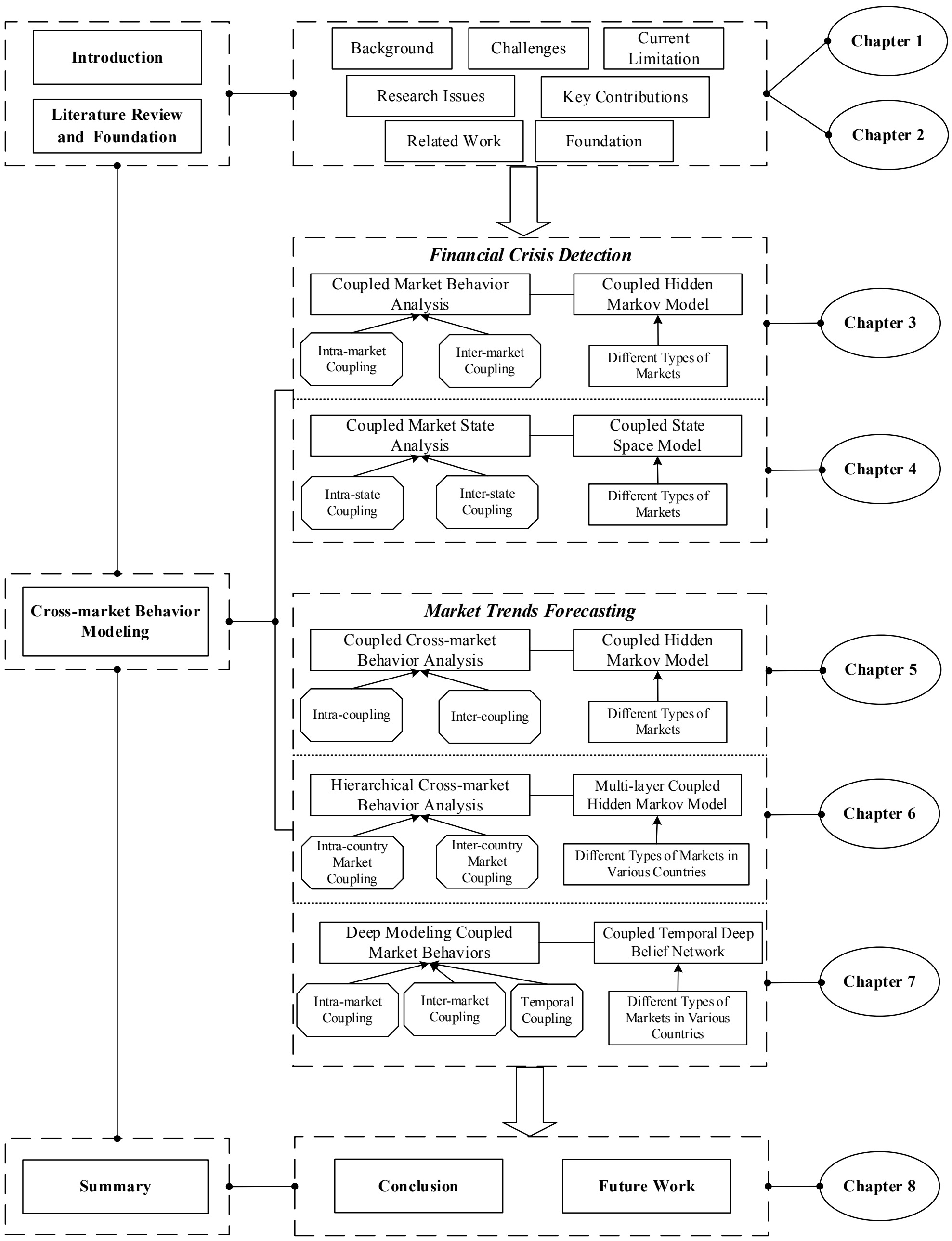


Figure 1.3: The Profile of Work in This Thesis

Chapter 2

Literature Review and Foundation

In this chapter, the related work and literatures are reviewed in terms of Coupled Behavior Analysis (CBA), Cross-market Analysis and three techniques related to this thesis. The understanding and corresponding foundation of CBA are introduced in Section 2.1. Two cross-market problems (i.e. financial crisis detection and market trend forecasting), corresponding approaches and their limitations are illustrated in Section 2.2. Section 2.3 focuses on the preliminaries of corresponding models and related literature which is needed for modeling the coupled cross-market behaviors in this thesis, including the Coupled Hidden Markov Model (CHMM), State Space Model (SSM) and Deep Belief Network (DBN).

2.1 Coupled Behavior Analysis

2.1.1 Behavior

Behavior is the range of actions and mannerisms made by individuals, organisms, systems, or artificial entities in conjunction with themselves or their environment, which includes the other systems or organisms around as well as the (inanimate) physical environment[1]. It is ubiquitous throughout our daily life. Generally speaking, behavior is the action or reaction of an entity, human or otherwise, to situations or stimuli in its environment (Cao 2010). In different applications and scenarios, behaviors exhibit different features. As illustrated in this thesis, the market trends (e.g. upward or downward) are behaviors conducted by actors: different markets.

Behavior is a key entity in understanding the driving forces and cause-effects of many issues (Cao 2010), especially for business problems. For example, it is widely agreed that a good understanding of customer behaviors is useful for boosting enterprise operations and enhancing business intelligence. An in-depth understanding of behaviors is increasingly recognized as an useful tool for exploring the underlying driving forces, causes and evolution of business problems, such as customer relationship management (Mozer, Wolniewicz, Grimes, Johnson & Kaushansky 2000), fraud detection (Fast, Friedland, Maier, Taylor, Jensen, Goldberg & Komoroske 2007), outlier detection (Hodge & Austin 2004) and multi-agent organizations (Dai, Zhang & Cao 2005).

2.1.2 Coupled Behavior

As illustrated above, behavior is an essential and critical activity which has been increasingly investigated in diverse fields including social (Pierce & Cheney 2013), engineering (Van Hemel, MacMillan, Zacharias et al. 2008), economics (Holcombe 1989) and computer science (Ilgen, Hulin, Association et al. 2000). However, this research mainly focuses on individual behaviors (Cao, Ou & Yu 2012). In practice, we realize that the behaviors are often coupled with each other. As illustrated in Figure 2.1, there are two financial markets (the stock market and currency market) at time $t-1$ and t. It is easy to understand that the stock market behaviors (here we can describe the market index as market behaviors) at time t not only depend on the history stock market behaviors at time $t-1$, but also influenced by currency market

[1]http://en.wikipedia.org/wiki/Behavior

behaviors at time $t-1$. This means that the behaviors of the two financial markets are coupled with each other. Thus, we need to pay more attention to the coupled behaviors which play a more fundamental role in social and business activities.

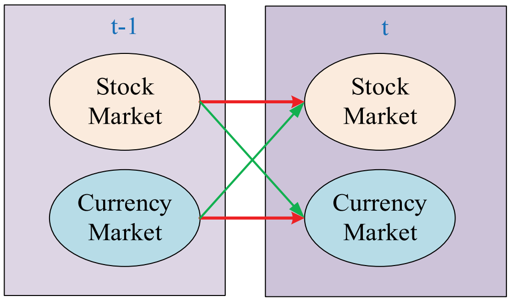


Figure 2.1: An Example of Coupled Behaviors between Financial Markets

To obtain an integrated understanding of coupled behavior, we introduce several definitions proposed in (Cao, Ou, Yu & Wei 2010, Cao et al. 2012).

Definition 2.1 (Behavior) *A behavior* $\mathbf{B}$ *is described as a four ingredient tuple* $\mathbf{B} = (\varepsilon, o, c, r)$*,*

- *Actor* ε *is the entity that issues a behavior or on which a behavior is imposed.*
- *Operation* o *is what an actor conducts in order to achieve certain goals.*
- *Context* c *is the environment in which a behavior takes place.*
- *Relationship* $r = \langle \theta(\cdot), \eta(\cdot) \rangle$ *is a tuple which reveals complex interactions within an actor's behaviors (named intra-coupled behaviors, represented by function* $\theta(\cdot)$*) and between multiple behaviors of different actors (inter-coupled behaviors, represented by function* $\eta(\cdot)$*).*

Suppose there are I actors, an actor ε_i undertakes J behaviors $\{\mathbf{B}_{i1}, \mathbf{B}_{i2}, \ldots, \mathbf{B}_{iJ}\}$. Actor ε_i's j^{th} behavior $\mathbf{B}_{ij}$ is a K-variable vector, its variable $[p_{ij}]_k$ reflects the k^{th} behavior property. Then a Behavior Feature Matrix is defined as follows:

$$FM(\mathbf{B}) = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \dots & \mathbf{B}_{1J} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \dots & \mathbf{B}_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{I1} & \mathbf{B}_{I2} & \dots & \mathbf{B}_{IJ} \end{pmatrix}$$

Then the intra-couplings are embodied through the relationships within one row of the above matrix, while the inter-couplings, namely the interactions between different actors are represented by the relationships of different rows of $FM(\mathbf{B})$.

Definition 2.2 (Coupled Behaviors) *Coupled behaviors* $\mathbf{BC}$ *refer to behaviors* $\mathbf{B}_{i_1j_1}$ *and* $\mathbf{B}_{i_2j_2}$ *that are coupled in terms of relationships* $f(\theta(\cdot), \eta(\cdot))$*, where* $(i_1 \neq i_2) \vee (j_1 \neq j_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J)$

$$\mathbf{BC} = (\mathbf{B}^{\theta}_{i_1j_1})^{\eta} * (\mathbf{B}^{\theta}_{i_2j_2})^{\eta} ::= \mathbf{B}_{ij}| \tag{2.1}$$
$$\sum_{i_1,i_2=1}^{I} \sum_{j_1,j_2=1}^{J_i} f(\theta_{j_1j_2}(\cdot), \eta_{i_1i_2}(\cdot)) \odot (\mathbf{B}_{i_1j_1}, \mathbf{B}_{i_2j_2})$$

where $\theta(\cdot)$ *is the coupling function indicating the actor* ε_i*'s behaviors* $\mathbf{B}_{ij}$ *are intra-coupled, actor* ε_i*'s behaviors* $\mathbf{B}_{ij}$ *are inter-coupled with each other in terms of the coupling function* $\eta(\cdot)$*, and* $f(\theta_{j_1,j_2}(\cdot), \eta_{i_1i_2}(\cdot))$ *is the coupling function denoting the corresponding relationships between* $\mathbf{B}_{i_1j_1}$ *and* $\mathbf{B}_{i_2j_2}$*,* $\sum_{i_1,i_2=1}^{I} \sum_{j_1,j_2=1}^{J_i} \odot$ *means the subsequent behaviors of* $\mathbf{B}$ *are* $\mathbf{B}_{i_1j_1}$ *coupled with* $f(\theta_{j_1}(\cdot), \eta_{i_1}(\cdot))$*,* $\mathbf{B}_{i_2j_2}$ *with* $f(\theta_{j_2}(\cdot), \eta_{i_2}(\cdot))$*.*

Definition 2.3 (Coupled Behavior Sequences) *Suppose* $\mathbf{BC}$ *is partitioned into* M *coupled behavior sequences, then*

$$\Phi(\mathbf{BC}) ::= \Phi(\mathbf{B})| \tag{2.2}$$
$$\sum_{t_1=1}^{T_1} \sum_{t_2=1}^{T_2} \cdots \sum_{t_M=1}^{T_M} f(\theta(\cdot), \eta(\cdot)) \odot \Phi_{\mathbf{B}_1\mathbf{B}_2 \dots \mathbf{B}_M}$$

where T_m *is the number of behavior instances for the* M^{th} *behavior sequence.* $f(\cdot)$ *illustrates the coupling relationships between two behavior sequences* $\Phi(\mathbf{B}_i)$

and $\Phi(\mathbf{B}_j)$*, which can be represented by* r_{ij} *which is a set for all* M *market behavior sequences* $(1 \leq i, j \leq M)$*. So if* $r_{ij} = \varnothing$*, there is no coupling relationship in behavior sequences* $\Phi(\mathbf{B}_i)$ *and* $\Phi(\mathbf{B}_j)$*, the bigger the* r_{ij}*, the stronger relationship between the two behavior sequences* $\Phi(\mathbf{B}_i)$ *and* $\Phi(\mathbf{B}_j)$*.*

Based on the above definitions, the coupled behavior analysis can be formalized as following (Cao et al. 2012):

Theorem 2.1 (Coupled Behavior Analysis (CBA)) *CBA is to build the objective function* $g(\mathbf{B})$ *under the condition that behaviors are coupled with each other by the coupling function* $f(\cdot)$*, and satisfy the following conditions:*

$$f(\cdot) ::= f(\theta(\cdot), \eta(\cdot)) \tag{2.3}$$

$$g(\mathbf{B}_{i_1 j_1})|f(\cdot) \geq g(\mathbf{B}_{i_2 j_2})|f(\cdot) \tag{2.4}$$

The vector-based representation of coupled behavior indicates the mapping from understanding the CBA problems to modeling them in real applications.

2.1.3 Related Techniques and Applications

There are many research outcomes that can be identified in the literature related to behavior analysis, including time series analysis (Yoon, Yang & Shahabi 2005), sequence analysis (Ayres, Flannick, Gehrke & Yiu 2002), interactive process modeling (Bickhard 2000) and behavior understanding (Cao 2010). However, the above research only focuses on individual sequences or behaviors, which cannot capture the coupled relationships between different behaviors from either the same, or different actors. To the best of our knowledge, there is very limited work that can be found in the literature which focuses on coupled behavior analysis. Some relevant approaches include:

- Multivariate Time Series Analysis. It is used to model and explain the interactions and co-movements among a group of time series and corresponding variables (Reinsel 2003). The researchers understand and model the multivariate time series from two kinds of perspectives: the statistical perspective and artificial intelligence perspective. Statistical methods include the Vector Auto-Regressive process (VAR) (Goebel, Roebroeck, Kim & Formisano 2003), Auto Regressive Moving Average (ARMA) (Hurvich & Tsai 1989), AutoRegressive Integrated Moving Average (ARIMA) (Box & Pierce 1970), Logistic regression (Hosmer Jr & Lemeshow 2004) and other non-linear approaches (Pole, West & Harrison 1994, Liu, Swift, Tucker, Cheng & Loizou 1999). Regarding the artificial intelligence methods, a clustering method based on calculating the degree of similarity between multivariate time series datasets is proposed in (Chandrakala & Sekhar 2008) for clustering multivariate time series, Tatavarty, Giridhar and Bhatnagar (Tatavarty, Bhatnagar & Young 2007) use Suffix Trees to discover all frequent patterns of multivariate time series in various dimensions. Other artificial intelligence applications include dependence detection of categorical data (Oates, Schmill, Cohen & Durfee 1999) and classification (Batal, Sacchi, Bellazzi & Hauskrecht 2009) also have been investigated in multivariate time series analysis.

- Sequence Analysis. Sequence analysis stems from bioinformatics and includes a very wide range of relevant research areas. There include many typical algorithms such as GSP (Srikant & Agrawal 1996), Spam (Ayres et al. 2002) and PrefixSpan (Pei, Han, Mortazavi-Asl, Pinto, Chen, Dayal & Hsu 2001). However, as illustrated in (Cao et al. 2012), they mainly focus on mining a single sequence, and this may lead to challenges when modeling multiple behavior sequences from different actors.

- Coupled Hidden Markov Model (CHMM). This is a model that was proposed to model multiple processes with coupling relationships (Oliver, Rosario & Pentland 2000), where one process is donated by one Hidden Markov Model (HMM) . In a CHMM (Saul & Jordan 1995, Brand

1997), the hidden variables (states) are assumed to interact with their neighbors, namely the state in one HMM at time t depends on the states of its own HMM and the states of other HMMs at time $t-1$. Since it is an important model in this thesis and also a useful tool in modeling multiple processes, we will discuss this model in detail in Section 2.3.1.

2.1.4 Discussion

In this section, we have provided the introduction of behavior, coupled behavior and coupled behavior analysis. The conclusions from above discussion can be listed as follows:

- Behavior plays a role as an internal driving force or cause for business appearance and associated problems. Cross-market problems (e.g. crisis detection) can be better investigated and learned if cross-market behaviors are understood and modeled correctly.
- In the real world, behaviors are coupled with each other in terms of some relationships. The coupled behaviors play a more fundamental role than individual behaviors in the driving force and effect of business problems. Based on this, this thesis focuses on modeling and learning the complex coupled behaviors for the purpose of cross-market analysis.
- Modeling and learning complex coupled behaviors is not a trivial task. It challenges the current approaches that are applied to cross-market problems. This is illustrated in Section 2.2.

2.2 Cross-market Analysis

Since correlations/linkages have been verified across different financial markets, much attention has been paid to cross-market analysis, which refers to run analysis on data across different markets to resolve real market problems. In this thesis, we mainly focus on the following two applications:

2.2.1 Financial Crisis Detection

Financial Crisis refers to the situations in which some financial assets suddenly lose a large part of their nominal value[2]. Since a crisis like the subprime mortgage crisis which began in 2007 has a large and damaging effect not only on individual investors but also on societies, there have been several attempts devoted to crisis detection in order to avoid big losses. Generally, the recent efforts at detecting financial crisis have taken the form of the following three related types.

A. Signal Approach

The Signal approach was proposed by Kaminsky and Reinhart in (Kaminsky & Reinhart 1999). The basic idea is to find the difference between economy behaviors on the eve of financial crises as opposed to normal periods. As illustrated in (Berg & Pattillo 1999, Alvarez-Plata & Schrooten 2004, Goldstein, Kaminsky & Reinhart 2000, Peng & Bajona 2008), market indexes such as the exchange rate or stock market index are often used as indicators. If they exceed a specified threshold, then a crisis signal will be produced. However, different market indicators behave differently when identifying a crisis (i.e. they may not produce a signal at the same time period). In order to combine all the information and various indicators from different markets at the same time, Kaminsky (Kaminsky 1999) proposes four methods to do information integration. But this does not solve all the problems, authors in (Yu, Wang, Lai & Wen 2010) illustrate that a very high noise-to-signals ratio will be produced if some of the indicators are strongly correlated, and the markets indexes are always closely related in the real world. The main limitation here is the approach relies on the selection of input variables (i.e. market indicators) and pays no attention to the interactions. This can lead to biased results.

[2] http://en.wikipedia.org/wiki/Financial_crisis

B. Linear Regression Approaches

The basic idea behind this kind of approach is to predict the probability of the occurrence of financial crisis for the following time period by using the historical data of some selected explanatory market variables (Sunderlin, Angelsen, Resosudarmo, Dermawan & Rianto 2001). The typical models are Logistic and Probit models. For instance, Kumar et al. adopt the Logistic approach to predict the emerging market currency crashes with pooled data on 32 developing countries from January 1985 to October 1999 (Kumar, Moorthy & Perraudin 2003). Similar work includes (Beckmann, Menkhoff & Sawischlewski 2006, Kalotychou & Staikouras 2006, Bussière & Fratzscher 2006). In addition, Eichengreen et al. (Eichengreen, Rose, Wyplosz, Dumas & Weber 1995) use the Probit approach to detect the exchange market crisis by using the data of 20 OECD countries from 1959 to 1993. Likewise, authors in (Berg & Pattillo 1999) apply a Probit regression technique to predict the Asia currency crisis.

The advantage of this kind of approach is that it can capture all the information contained in the selected market variables (Lin, Khan, Chang & Wang 2008). However, the occurrence of financial crisis is a rare event and as it reveals non-linear characteristics the models with linear assumptions may lead to disappointing results. Furthermore, it is difficult to collect the large amount of data required from various market indicators.

C. Model-based Approaches

As the computational technology is widely used in business prediction, model-based approaches have begun to develop. This kind of approach adopts artificial intelligence and machine learning techniques to provide financial crisis detection (Yu et al. 2010). Techniques such as Neural Network (NN) (Kim, Hwang & Lee 2004), Support Vector Machine (SVM) (Shin, Lee & Kim 2005), Fuzzy Logic (FL) (Lin et al. 2008) and Decision Tree (DT) are adopted by researchers. Some recent work reveals that the Artificial Neural Network (ANN) is a useful tool in crisis detection and it provides promising results. For example, by using the stock market variables, authors in (Kim, Oh, Sohn & Hwang 2004) illustrate the competitive nature of ANN in building early warning systems for crisis detection as compared to other data mining techniques. Celik and Karatepe (Celik & Karatepe 2007) test ANN in banking crisis forecasting by using data with the same date and cross sectional data, and the results reveal that ANN is effective in evaluating and forecasting bank crisis. Similar work can be found in (Ozkan-Gunay & Ozkan 2007, Baesens, Setiono, Mues & Vanthienen 2003, Atiya 2001). However, the ANN also has limitations: the black-box nature leads to difficulty in searching the relationships between the indicators. Future, the results rely on the selection of indicators, but in the real world the data for some indicators is hard to obtain.

2.2.2 Market Trend Forecasting

Financial market trend forecasting mainly focuses on developing approaches that can successfully predict index trends, and the aim of the predictions to help investors obtain high profits. The literature addressing this area of market trend forecasting can be categorized into the following two types.

A. Time Series Models

Time series analysis assumes that the current values of a time series have a functional relationship with past values, and it uses historical data to infer future trend behaviors. It includes many classical approaches such as Linear and Multi-linear Regression, Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) models. Typical representatives are ARIMA (Box, Jenkins & Reinsel 2013) and Logistic models in linear regression. The ARIMA model is used in (Contreras et al. 2003) to analyze time series from the mainland Spain and California markets to predict the next-day's electricity market price. Authors in (Tseng, Tzeng, Yu & Yuan 2001) develop a fuzzy ARIMA model and apply it to predict the NT (New Taiwan Dollar) dollars to US dollars, with 40 observations from

1 August 1996 to 16 September 1996. The Multinomial Logistic Regression is applied in (Upadhyay, Bandyopadhyay & Dutta 2012) to predict the outperforming stock in the Indian market. Similar research can be found in (Quagrainie 2004, Laitinen & Laitinen 2001). In (Chen & Chen 2011), a fuzzy multivariate time series analysis is used to forecast the daily Taiwan stock index. In addition, other models like Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) models (Marcucci 2005) are applied in forecasting financial market trend.

However, this kind of method suffers from market changes of the corresponding input market variables, as they make predictions highly dependent on historical observations and input variables. Very limited work can be found that addresses the underlying complex interactions between market indicator series, which fundamentally drive cross-market movements.

B. Machine Learning-based Models

With the development of AI, machine learning-based models are increasingly explored for financial market forecasting. ANN represents one widely used machine learning method for market trend forecasting. It is first used in (White 1988) to predict the IBM daily stock price movement. Authors in (Huarng & Yu 2006) forecast the Taiwan stock index for the years 1991-2003 by using a backpropagation neural network, and the results display the effectiveness of ANN. In (Halliday 2004), the authors apply ANN to predict the trends of individual equities listed in the New York Stock Exchange from 1985 to 2000. Similar work includes (Olson & Mossman 2003, Pan et al. 2005, Hyup Roh 2007). Another popularly used method is HMM which checks any systematic patterns in time series for prediction. Authors in (Weigend & Shi 1998) show promising result while using HMM to predict the changes of the daily S&P500 index. In other work, Hassan and Nath (Hassan & Nath 2005) apply HMM to generate a one-day forecast of stock movement for interrelated markets. In addition, other machine learning-based methods used in forecasting financial markets include SVM (Cao &

Tay 2001), fuzzy logic (Simutis 2000) and hybrid models (Hassan, Nath & Kirley 2007). Although these methods explore couplings in a market or the correlations between markets, they do not effectively address the intra- and inter-market couplings between various financial markets with multiple distinct financial indicators. Recently, the Coupled Hidden Markov Model (CHMM) was deployed to learn the coupled market behaviors (Cao et al. 2012) and this captures the hidden couplings in the multiple time series.

2.2.3 Discussion

In this section, we present a literature review which focuses on financial crisis detection and market trend forecasting. The conclusions from the literature review include but are not limited to:

- Current cross-market modeling relies on the selection of input market variables, while pays no attention to the complex interactions between markets. Limited work emphasizes the fact that complex interactions play a key role in the internal driving force of cross-market movements.
- Complex coupling relationships across markets, either in terms of an intra- or inter- perspective, are often ignored or only weekly addressed. Effective approaches for modeling the complex coupling relations are not available, since the existing methods mainly focus on simple correlations.
- The existing work often analyzes the cross-market directly from observation/data and overlooks the hidden features behind the data, which may in turn lead to biased results.
- Some assumptions under the approaches are not valid. For example, it is not proper to use linear assumption since both the financial crisis and the market reveal non-linear characteristics.

The above summary is also consistent with the limitations listed in Section 1.2.2. To overcome the limitations and address the challenges, this thesis seeks to deep understand, model and learn the complex cross-market behaviors, building on the classic methods which are canvassed in Section 2.3.

2.3 Related Techniques

2.3.1 Coupled Hidden Markov Model

A. Markov Model

A Markov Model is a stochastic model that assumes the Markov property[3] and it was named by Andrei Markov in the early twentieth century (Markov 2006). It is often used to train sequential data and it models a process where the hidden state in the data sequence depends on previous states. As shown in Figure 2.2, if we consider a system with a set of H distinct states $\{Z_1, Z_2, \ldots, Z_H\}$ (here H equals to 3).

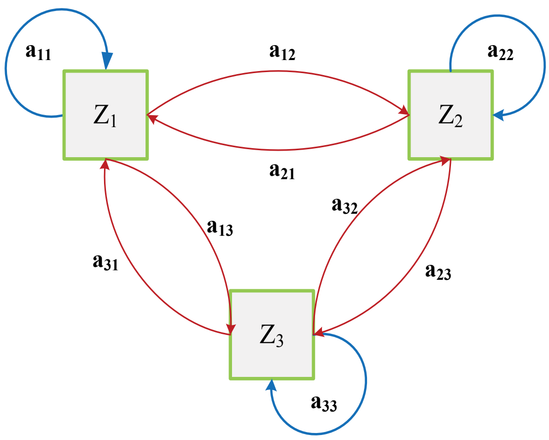


Figure 2.2: A Markov Chain with Three States

At each time slot t (for example, $t = 1, 2, \ldots$), the system conducts a state change according to a set of state transition probabilities. In order to fully describe the system from a probabilistic perspective, the current state (e.g. at time t) and previous states should be specified. In some real

[3]http://en.wikipedia.org/wiki/Markov_model

applications, a first order Markov process is often adopted (Elliott, Aggoun & Moore 1994). The first order Markov process means the current state at time t only depends on the state at time $t-1$, and has nothing to do with the whole history states. According to (Rabiner & Juang 1986), the first order Markov properties can be described as follows:

$$P(z_{t+1} = Z_j \mid z_t = Z_i, z_{t-1} = Z_j, \dots) = P(z_{t+1} = Z_j \mid z_t = Z_i) \qquad (2.5)$$

where z_{t+1} is the state at time $t+1$. Since the right hand of Equation (2.5) is independent of time, then the set of transition probabilities a_{ij} is given by:

$$a_{ij} = P(z_t = Z_j \mid z_{t-1} = Z_i), 1 \leq i, j \leq H \qquad (2.6)$$

where a_{ij} has the following properties since it is a probability:

$$a_{ij} \geq 0, \quad \sum_{j=1}^{N} a_{ij} = 1 \qquad (2.7)$$

The above process can be called an observed Markov model if the output is the state sequence and each state is connected with an observation (Rabiner 1989).

B. Hidden Markov Model

Since the the Markov model introduced above has many limitations in real applications, we import the Hidden Markov model (HMM) where the observation is a probabilistic function of the state (Sonnhammer, Von Heijne, Krogh et al. 1998). HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states[4]. The mathematical theory behind the HMM was developed by L.E. Baum and his coworkers (Baum & Petrie 1966, Baum, Eagon et al. 1967, Baum, Petrie, Soules & Weiss 1970).

a. Elements of HMM

A first order HMM model is composed of the following elements (Rabiner 1989):

[4]http://en.wikipedia.org/wiki/Hidden_Markov_model

- A set of hidden states $\{Z_1, Z_2, \ldots, Z_H\}$, where H is the number of hidden states in the HMM.
- A set of observation symbols $\{X_1, X_2, \ldots, X_V\}$, where V is the number of observation symbols.
- Initial state probability distribution $\pi = \{\pi(i)\}$, where for $1 \leq i \leq H$

$$\pi(i) = P(z_1 = Z_i), \quad s.t. \sum_{i=1}^{H} \pi_i = 1 \tag{2.8}$$

- State transition probability matrix $A = \{a_{ij}\}$, where for $1 \leq i, j \leq H$

$$a_{ij} = P(z_t = Z_j \mid z_{t-1} = Z_i), \quad s.t. \sum_{j=1}^{H} a_{ij} = 1 \tag{2.9}$$

- Observation probability matrix $B = \{b_j(v)\}$,where for $1 \leq j \leq H$

$$b_j(v) = P(o_t = X_v \mid z_t = Z_j), \quad s.t. \sum_{v=1}^{V} b_j(v) = 1 \tag{2.10}$$

For convenience, we denote a standard HMM as $\lambda = (\pi, A, B)$.

b. Three Basic Problems of HMM

Given the form of HMM, there are three fundamental problems of interest in terms of HMM:

- Problem 1 (Observation Evaluation Problem): Given a model $\lambda = (\pi, A, B)$ and the observation sequence $O = \{O_1, O_2, \ldots, O_T\}$, how to compute the probability of an observation sequence given the model, $P(O \mid \lambda)$?
- Problem 2 (Decoding Problem): Given a model $\lambda = (\pi, A, B)$ and the observation sequence $O = \{O_1, O_2, \ldots, O_T\}$, how to find the corresponding state, $argmax_Z P(Z \mid O, \lambda)$?
- Problem 3 (Parameter Learning Problem): Given the observation sequence $O = \{O_1, O_2, \ldots, O_T\}$, how to choose the model parameters to maximize $P(O \mid \lambda)$?

To solve problem 1, efficient algorithms are the well-known forward/backward procedures (Baum & Sell 1968). The problem 2 is often solved through the Viterbi algorithm (Viterbi 1967, Shinghal & Toussaint 1979). Problem 3 is the most important and difficult problem, and the main training algorithm is the Baum-Welch algorithm, first described and proved to converge in (Baum & Petrie 1966, Baum et al. 1967). It is also known as a case of the Expectation-Maximization (EM) algorithm (Dempster, Laird & Rubin 1977). In the following, we mainly discuss solutions to problems 1 and 3 since they are more meaningful for this thesis.

Solution to Problem 1: The Forward-backward Procedure

As illustrated in (Rabiner 1989), the observation evaluation problem $P(O \mid \lambda)$ can be written as:

$$P(O \mid \lambda) = \sum P(O \mid Z, \lambda) P(Z \mid \lambda) \tag{2.11}$$
$$= \sum_{z_1, z_1, \ldots, z_T} \pi_{z_1} b_{z_1}(o_1) a_{z_1 z_2} b_{z_2}(o_2) \ldots a_{z_{T-1} z_T} b_{z_T}(o_T)$$

where $\{z_1, z_2, \ldots, z_T\}$ is the hidden state sequence. Then the problem can be solved using the forward/backward procedure, where the forward variable is defined as

$$\alpha_t(i) = P(o_1, o_2, \ldots, o_t, Z_{i,t} \mid \lambda) \tag{2.12}$$

and this can be calculated inductively as follows:

1. Initialization

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq H \tag{2.13}$$

2. Induction

$$\alpha_t(j) = b_j(o_{t+1}) \sum_{j=1}^{H} a_{ij} \alpha_{t-1}(i), \quad 2 \leq t \leq T, 1 \leq j \leq H \tag{2.14}$$

3. Termination

$$P(O \mid \lambda) = \sum_{i=1}^{H} \alpha_T(i) \tag{2.15}$$

The backward variable calculation applies a similar way. For more details refer to (Rabiner 1989).

Solution to Problem 3: The Baum-Welch Training Algorithm (EM)

The parameter problem is the most important and difficult problem in HMM. There is no known analytic way to find the parameters $\lambda = (\pi, A, B)$ to maximize the probability of observation sequence $P(O \mid \lambda)$. However, we can use the Baum-Welch algorithm (Baum & Petrie 1966, Baum et al. 1967) by which the $P(O \mid \lambda)$ is locally maximized with the iterative process. And the re-estimation formulas are as follows:

- State transition probability

$$\begin{aligned} \bar{a}_{ij} &= \frac{expected\, number\, of\, transitions\, form\, state Z_i\, to\, Z_j}{expected\, number\, of\, transitions\, form\, state\, Z_i} \\ &= \frac{a_{ij}(\partial P(O|\lambda)/\partial a_{ij})}{\sum_k a_{ik}(\partial P(O|\lambda)/\partial a_{ik})} \end{aligned} \tag{2.16}$$

- Observation probability

$$\begin{aligned} \bar{b}_j(v) &= \frac{expected\, number\, of\, transitions\, form\, observation X_v}{expected\, number\, of\, times\, in\, state\, j} \\ &= \frac{b_j(v)(\partial P(O|\lambda)/\partial b_j(v))}{\sum_l b_j(l)(\partial P(O|\lambda)/\partial b_j(l))} \end{aligned} \tag{2.17}$$

- Prior probability

$$\begin{aligned} \bar{\pi}_i &= expected\, number\, of\, times\, at\, state\, i\, at\, time\, t = 1 \\ &= \frac{\pi_i(\partial P(O|\lambda)/\partial \pi_i)}{\sum_k \pi_k(\partial P(O|\lambda)/\partial \pi_k)} \end{aligned} \tag{2.18}$$

c. Application

HMM is popular owing to its expressive power in modeling real world problems with low computation complexity. It has found success in several areas, including speech recognition (Rabiner 1989, Varga & Moore 1990), bioinformatics (Baldi & Brunak 2001, Söding 2005), video analysis (Boreczky & Wilcox 1998, Töreyin, Dedeoğlu & Çetin 2005) and time sequence analysis

(Hassan & Nath 2005, Hassan et al. 2007). In addition, there are a variety of new HMM architectures which have been proposed to address real world problems and to overcome the limitations of the standard HMM, including hierarchical HMM (Fine, Singer & Tishby 1998, Nguyen, Phung, Venkatesh & Bui 2005), coupled HMM (Zhong & Ghosh 2002, Oliver et al. 2000, Rezek, Gibbs & Roberts 2002), factorial HMM (Ghahramani & Jordan 1997, Chen, Liang, Zhao, Hu & Tian 2009) and input-ouput HMM(Bengio & Frasconi 1996).

In the section below, we introduce the extended Coupled Hidden Markov Model (CHMM) which is of particular relevance to this thesis.

C. Extended Coupled Hidden Markov Model

CHMM (Oliver et al. 2000) is a model that was proposed to model multiple processes with coupling relationships. CHMM consists of more than one chain of HMMs, and each HMM represents one process. In CHMM, the state of any chain of HMM at time t depends on not only the states of its own chain but also the states of other chains at time $t-1$. These are namely the interactions between the processes.

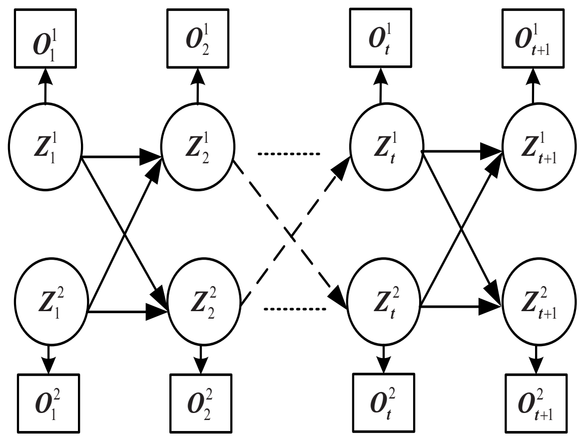


Figure 2.3: A CHMM with Two Chains

Figure 2.3 is a typical standard CHMM with two chains. Suppose there are C HMMs coupled together, then the state transition probability is $P(Z_t^c \mid Z_{t-1}^1, Z_{t-1}^2, \ldots, Z_{t-1}^C)$ which is different from HMM with $P(Z_t^c \mid Z_{t-1}^c)$. In other words, the state transition probability in CHMM is much more complicated than HMM since it is described by a $C+1$ dimensional matrix, and the number of parameters here is H^C (H is the number of hidden states for all HMMs). To resolve the issue, Brand (Brand 1997) computes the complex joint state transition probability as the product of various marginal conditional probabilities:

$$P(Z_t^c \mid Z_{t-1}^1, Z_{t-1}^2, \ldots, Z_{t-1}^C) = \prod_{c'=1}^{C} P(Z_t^c \mid Z_{t-1}^{c'}) \tag{2.19}$$

Then the transition probability parameter space is reduced (Brand, Oliver & Pentland 1997). However, as illustrated in (Zhong & Ghosh 2002), Equation (2.19) does not hold since the right hand side does not equal to one. In this thesis, we follow the method proposed by Zhong in (Zhong & Ghosh 2001), which models the joint transition probability as:

$$P(Z_t^c \mid Z_{t-1}^1, Z_{t-1}^2, \ldots, Z_{t-1}^C) = \sum_{c'=1}^{C} (r_{c'c} P(Z_t^c \mid Z_{t-1}^{c'})) \tag{2.20}$$

where the joint transition probability is modeled as a linear combination of various marginal conditional probabilities. Here $r_{c'c}$ is the coupling coefficient which evaluates the coupling weight from model c' to model c, the bigger the $r_{c'c}$, the more $Z_{t-1}^{c'}$ affects Z_t^c.

a. Elements of CHMM

Suppose there are C coupled HMMs, H is the number of states of the Markov chains, $\{Z_1, Z_2, \ldots, Z_H\}$ is a set of hidden states, where z_t is the hidden state at time t. V is the number of observation symbols, $\{X_1, X_2, \ldots, X_V\}$ is a set of observation symbols, $O = \{O_1, O_2, \ldots, O_T\}$ is an observation sequence, o_t is the observation at time t. Then the corresponding elements of a CHMM can be defined as follows:

- The number of Markov chains (they are coupled with each other) C.

- A set of hidden states $\{Z_1, Z_2, \ldots, Z_H\}$, where H is the number of hidden states in the HMM.
- A set of observation symbols $\{X_1, X_2, \ldots, X_V\}$, where V is the number of observation symbols.
- Initial state probability distribution $\pi = \{\pi(i^{(c)})\}$, where for $1 \leq i \leq H^{(c)}$

$$\pi(i^{(c)}) = P(z_1^c = Z_i), \quad s.t. \sum_{i=1}^{H^{(c)}} \pi_i = 1 \tag{2.21}$$

- State transition probability matrix $A = \{a_{ij}^{(c',c)}\}$, where for $1 \leq c', c \leq C, 1 \leq i \leq H^{(c')}, 1 \leq j \leq H^{(c)}$

$$a_{ij}^{(c',c)} = P(z_t^c = Z_j \mid z_{t-1}^{c'} = Z_i), \quad s.t. \sum_{j=1}^{H^{(c)}} a_{ij}^{(c',c)} = 1 \tag{2.22}$$

- Observation probability matrix $B = \{b_j^{(c)}(v)\}$, where for $1 \leq j \leq H^{(c)}$

$$b_j^{(c)}(v) = P(o_t^{(c)} = X_v \mid z_t^{(c)} = Z_j), \quad s.t. \sum_{v=1}^{V} b_j^{(c)}(v) = 1 \tag{2.23}$$

- Coupling coefficient $R = \{r_{c'c}\}$, where for $1 \leq c, c' \leq C$

$$s.t. \sum_{c'=1}^{C} r_{c'c} = 1 \tag{2.24}$$

For convenience, we refer to the complete set of parameters of a CHMM as $\Omega = (A, B, R, \pi)$.

b. Extended Forward-backward Procedure

Here we introduce the extended forward-backward procedure that is used to solve $P(O \mid \Omega)$ (problem 1 in three basic problems of HMM). Note there are C coupled HMMs, and each o_t is a vector $(o_t^{(1)}, o_t^{(2)}, \ldots, o_t^{(C)})^T$, then the forward variable is defined as (Zhong & Ghosh 2001):

$$\alpha_t(i_1, i_2, \ldots, i_C) = P(o_1, o_2, \ldots, o_t, Z_{t,i_1}, Z_{t,i_2}, \ldots, Z_{t,i_C} \mid \Omega) \tag{2.25}$$

The forward variable can be calculated through following inductive processes:

1. Initialization

$$\alpha_1^{(c)}(i) = \pi_i^{(c)} b_i^{(c)}(o_1^{(c)}), \quad 1 \leq i \leq H^{(c)} \tag{2.26}$$

2. Induction

$$\alpha_t^{(c)}(j) = b_j^{(c)}(o_t) \sum_{c'}^{C} \theta_{c'c} \sum_{i=1}^{H^{(c')}} \alpha_{t-1}(i) a_{ij}^{(c',c)}, \quad 2 \leq t \leq T, 1 \leq j \leq H^{(c)} \tag{2.27}$$

3. Termination

$$P(O|\Omega) = \prod_{c=1}^{C} P^{(c)} = \prod_{c=1}^{C} (\sum_{j}^{H^{(c)}} \alpha_T^{(c)}(j)) \tag{2.28}$$

c. Parameter Estimation

From the above procedure, it can be seen that each parameter $\omega \in \Omega$ is subject to stochastic constraint, i.e. the sum equals to one. This leads us to work out the parameter learning by constrained optimization techniques. Here the parameter estimation is solved by the classical method of Lagrange multipliers illustrated in (Zhong & Ghosh 2001). Here we use the state transition probability $A = a_{ij}^{(c'c)}$ as an example since all parameters in Ω are subject to similar constraints. Let $\mathcal{L}$ be the lagrangian of $P(O|\Omega)$, and then

$$\mathcal{L} = P(O|\Omega) + \sum_{c',c} \lambda^{(c',c)} (\sum_{j=1}^{H} a_{ij}^{(c',c)} - 1) \tag{2.29}$$

where the $\lambda^{c',c}$ are the Lagrange multipliers. It can be verified that $P(O|\Omega)$ is locally maximized when

$$\bar{a}_{ij}^{(c',c)} = \frac{a_{ij}^{(c',c)} \partial P(O|\Omega) / \partial a_{ij}^{(c',c)}}{\sum_{k=1}^{H^{(c)}} a_{ik}^{(c',c)} \partial P(O|\Omega) / \partial a_{ik}^{(c',c)}} \tag{2.30}$$

where $\partial P(O|\Omega)/\partial a_{ij}^{(c',c)} = \sum_c (\frac{P(O|\Omega)}{\partial \alpha_T^{(c)}(j)} \sum_j \frac{\partial \alpha_T^{(c)}(j)}{\partial a_{ij}^{(c',c)}})$ Following a similar process, we can obtain the re-estimation formulas for other parameters as follows:

- Observation probability

$$\bar{b}_j^{(c)}(v) = \frac{b_j^{(c)}(v)\partial P(O|\Omega)/\partial b_j^{(c)}(v)}{\sum_l b_j^{(c)}(l)\partial P(O|\Omega)/\partial b_j^{(c)}(l)} \tag{2.31}$$

where $\partial P(O|\Omega)/\partial b_j^{(c)}(v) = \sum_c(\frac{P(O|\Omega)}{\partial \alpha_T^{(c)}(j)} \sum_j \frac{\partial \alpha_T^{(c)}(j)}{\partial b_j^{(c)}(v)})$

- Prior probability

$$\bar{\pi}_j^{(c)} = \frac{\pi_j^{(c)}\partial P(O|\Omega)/\partial \pi_j^{(c)}}{\sum_k \pi_k^{(c)}\partial P(O|\Omega)/\partial \pi_k^{(c)}} \tag{2.32}$$

where $\partial P(O|\Omega)/\partial \pi_j^{(c)} = \sum_c(\frac{P(O|\Omega)}{\partial \alpha_T^{(c)}(j)} \sum_j \frac{\partial \alpha_T^{(c)}(j)}{\partial \pi_j^{(c)}})$

- Coupling coefficient

$$\bar{r}_{c'c} = \frac{r_{c'c}\partial P(O|\Omega)/\partial r_{c'c}}{\sum_u r_{u'c}\partial P(O|\Omega)/\partial r_{u'c}} \tag{2.33}$$

where $\partial P(O|\Omega)/\partial r_{c'c} = \sum_c(\frac{P(O|\Omega)}{\partial \alpha_T^{(c)}(j)} \sum_j \frac{\partial \alpha_T^{(c)}(j)}{\partial r_{c'c}})$

D. Applications

CHMM is capable of modeling multiple interactive sequences, and the application of CHMM including action recognition (Brand 1997, Natarajan & Nevatia 2007), time series interactions (Rezek & Roberts 2000), bioinformatics (e.g. disease interactions) (Sherlock, Xifara, Telfer & Begon 2013). All these lead to the consideration that CHMM is an appropriate tool to model and analyze the complex coupled behaviors illustrated in Section 2.1.

2.3.2 State Space Model

State space model (SSM) stems from system theory and the term "state space" originated in the area of control engineering (Kalman 1960). SSM refers to a class of a probabilistic graphical model (Koller & Friedman 2009) that represents the probabilistic dependence between the latent state variable and the observations. In SSM, the information of hidden states is conditioned on their previous hidden states which obey the Markov property illustrated in Section 2.3.1. In addition, given the hidden states at one time step, the observations at that time step are statistically independent from other observations.

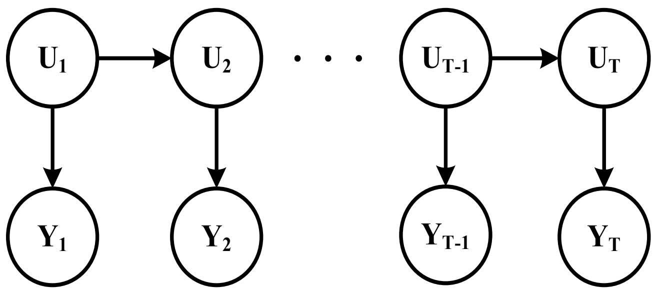


Figure 2.4: An Example of State Space Model

As illustrated in Figure 2.4, where $\{U_1, U_2, \ldots, U_T\}$ is the hidden state vector, $\{Y_1, Y_2, \ldots, Y_T\}$ is a real valued observation vector and it is generated from the hidden state vector. It is easy to see that the output Y_t is conditionally independent from all other variables given the state U_t, and if given U_{t-1}, the hidden state U_t is conditionally independent from $U_1, U_2, \ldots, U_{t-2}$. The joint probability for the sequences of states and observations can therefore be depicted as follows:

$$P(U_t, Y_t) = P(U_1)P(Y_1 \mid U_1)\prod_{t=2}^{T} P(U_t \mid U_{t-1})P(Y_t \mid U_t) \tag{2.34}$$

The most general SSMs are defined by the following two equations:

$$ObservationEquation \;\; Y_t = H_t U_t + v_t \tag{2.35}$$

$$StateEquation \;\; U_t = F_t U_{t-1} + w_t \tag{2.36}$$

where Y_t is called the *observation vector*, and U_t is the *state vector*, H_t links the state vector to the observations, F_t is a *state transition matrix* and v_t, w_t are the *control vectors*.

A. Inference

Probabilistic inference is the process of computing the posterior distribution of some variables given other variables (usually observed). The inference problem in SSM is to estimate the posterior distributions of hidden variables while giving the model parameters and an observation sequence. There are three typical cases in inference problem which are often considered (Goodwin & Sin 2013, Ghahramani & Hinton 2000).

- Filtering : Given the observation sequence $\{Y_1, Y_2, \ldots, Y_t\}$, how to compute the probability of current hidden variable U_t $P(U_t \mid Y_1, Y_2, \ldots, Y_t)$?
- Smoothing: Given the observation sequence $\{O_1, O_2, \ldots, O_T\}$, how to compute the probability of hidden variable at time t U_t $P(Z_t \mid O_1, O_2, \ldots, O_T)$ (here $T > t$)?
- Prediction: Given the observation sequence $\{Y_1, Y_2, \ldots, Y_t\}$, how to compute the probabilities of future hidden variables and observations, for example, $P(U_{t+1} \mid Y_1, Y_2, \ldots, Y_t)$ or $P(Y_{t+1} \mid Y_1, Y_2, \ldots, Y_t)$?

The first filtering problem can be solved by the well-known recursive algorithm named the Kalman filter. The filter was proposed by Kalman and Bucy in (Kalman & Bucy 1961). The algorithm is very similar with the forward procedure in HMM. The combined forward and backward recursive algorithms are applied to the smoothing problem, where the backward recursion is analogous to the backward procedure in HMM. The Kalman filter is used in the forward direction to compute the probability of U_t given $\{Y_1, Y_2, \ldots, Y_t\}$. A backward recursion then is used to complete the computation from time T to t (Rauch 1963, Haykin, Haykin & Haykin 2001). For the prediction problem, the solution is as follows: first compute the probability of hidden variable in time t $P(U_t \mid Y_1, Y_2, \ldots, Y_t)$, then the probabilities of the hidden variable and observation at a future time can be simulated in the forward direction following Equations (2.35) and (2.36).

B. Parameter Estimation

The parameter estimation problem of a state space model is known as the system identification problem. An often used method is the EM algorithm proposed in (Dempster et al. 1977). Given the observations, the E-step is used to compute the posterior probabilities of hidden variables and fix the current parameters while the M-step is used to maximize the expected log likelihood of the parameters by using the results in the E-step (Ghahramani & Hinton 1996, Murphy 2002).

From the above we can see that the HMM and SSM are very similar and the main difference is that HMM has discrete hidden states while SSM has continuous states. It is realized that HMMs are discrete non-linear SSMs; or equivalently linear SSMs are linear continuous HMMs (Ghahramani & Hinton 2000). In addition, there are some hybrid models which combine the linear dynamic state space model with the discrete transition structure of HMM (Kim 1994, Frühwirth-Schnatter 2006).

C. Applications

State-space models as an important mathematical tool have been widely used in many different fields. Numerous applications can be found in the finance area, including modeling stationary time series (Durbin & Koopman 2012), Nonstationarity in Mean (Chatfield 2013), Nonstationarity in Variance (Godsill, Doucet & West 2000) and Signal Extraction (Kitagawa 1998). It can also be found in other applications such as ecology (Patterson, Thomas, Wilcox, Ovaskainen & Matthiopoulos 2008), transportation (Stolyar 2004) and neuroscience.

2.3.3 Deep Belief Networks

A. Deep Architecture

The literature on neural networks shows that the Artificial Neural Network (ANN) is a commonly-used tool in Artificial Intelligence (AI) (e.g. the vision, language and finance areas). However, theoretical results suggest that single-layered neural networks cannot express complicated functions which represent high-level abstractions in some real applications (Bengio 2009). Then deep architectures which are composed of multiple levels of operations are introduced to alleviate the problem.

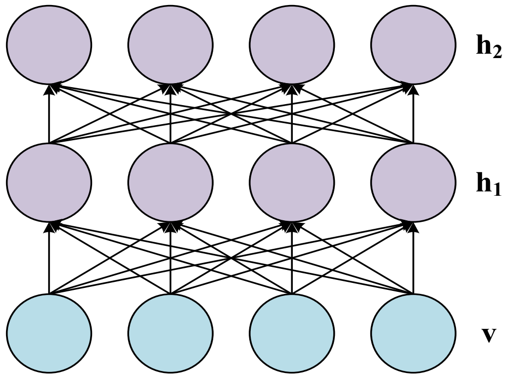


Figure 2.5: Three Layer Neural Network with Two Hidden Layers. The blue units represent the visible layer while the purple units represent hidden layers.

Deep architectures refer to the multi-layer networks which are compositions of many layers of adaptive non-linear components (Bengio & LeCun 2007, Mo 2012), where each layer corresponds to a different area of cortex, and the number of layers denote the depth of architecture. A deep architecture then is a layered architecture that has many levels. The bottom layer is the visible layer which represents the observations, the higher layers are all hidden layers which contain latent variables (features) that cannot be observed directly from the data. Figure 2.5 is a typical three layer neural network with two hidden layers.

Generally speaking, deep architectures have the following advantages:

- Building a hierarchical architecture of data (observations). Each subsequent hidden layer represents the new and higher features from the data by performing a transition from the previous hidden layer. Therefore, the deep architecture can learn abstract features automatically from the data and this removes the vulnerability of human-selected features. This is very important since we are living in a big data age and the amount of data continues to grow exponentially.

- Performing non-local generalization (Bengio 2009). Local generalization is commonly-used by some shallow architecture models (e.g. support vector machines), however, it is sometimes unsuitable since it poorly characterizes the behavior of highly varying functions.

There are several deep architectures. Here we introduce a widely used type, Deep Belief Networks (DBN), which is related to this thesis. DBN is a generative graphical model, or a type of deep neural network composed of data variables in the first layer (visible layer) and multiple layers of latent variables ("hidden units") above it. Further, there are connections between the hidden layers but not between the units within each layer[5]. Figure 2.5 is a specific example. We introduce Restricted Boltzmann Machines (RBMs) in the following section as it is the building block of DBN.

B. Restricted Boltzmann Machines

Boltzmann Machine was proposed as a type of stochastic recurrent neural network in 1985 (Ackley, Hinton & Sejnowski 1985). It is often used to learn important aspects of an unknown probability distribution based on samples from the distribution. However, this learning process is difficult and time-consuming in real world applications. In order to simplify the computation, restricted Boltzmann machines were created.

As shown in Figure 2.6, an RBM often consists of two layers of variables, a visible layer (the blue units $\mathbf{v}$) and a hidden layer (the purple units $\mathbf{h}$). Here we treat all units ($\mathbf{v}$ and $\mathbf{h}$) as binary random variables. It can be easily seen from the figure that there exist transformations between hidden units and visible units, namely each visible unit is a transformation of hidden units, and vice versa. In an RBM, the hidden units $\mathbf{h}$ are conditionally independent

[5] http://en.wikipedia.org/wiki/Deep_belief_network

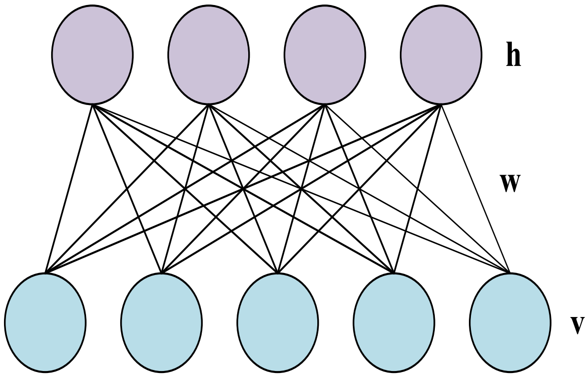


Figure 2.6: An Example of Restricted Boltzmann Machine

given the visible units $\mathbf{v}$ and vice versa. In addition, the connections are only between $\mathbf{h}$ and $\mathbf{v}$. The joint density of the distribution represented by an RBM is defined through an energy-based function:

$$P(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = exp(-E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}))/Z(\boldsymbol{\theta}) \tag{2.37}$$

where Z is a normalization constant named the partition function (Taylor 2009). And $E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})$ is an energy function:

$$E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = -\mathbf{v}^{\mathbf{T}}\mathbf{W}\mathbf{h} - -\mathbf{a}^{\mathbf{T}}\mathbf{v} - \mathbf{b}^{\mathbf{T}}\mathbf{h} \tag{2.38}$$

where $\mathbf{v} \in \{0, 1\}^D$ is a vector of binary visible units and $\mathbf{h} \in \{0, 1\}^F$ is a vector of binary hidden units. $\mathbf{W} \in \mathbb{R}^{D \times F}$ encodes the interactions between visible units $\mathbf{v}$ and hidden units $\mathbf{h}$. $\mathbf{a} \in \mathbb{R}^D$ and $\mathbf{b} \in \mathbb{R}^F$ denote the biases of $\mathbf{v}$ and $\mathbf{h}$ separately. Hence, $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{a}, \mathbf{b}\}$ are the model parameters that need to learn.

Due to the factorization trick introduced in (Bengio, Courville & Vincent 2013) and the linear form of the energy function, the conditional distribution w.r.t. visible units $\mathbf{v}$ and hidden units $\mathbf{h}$ can be factored as follows:

$$P(\mathbf{v} \mid \mathbf{h}) = \prod_i P(v_i \mid \mathbf{h}) \tag{2.39}$$

$$P(\mathbf{h} \mid \mathbf{v}) = \prod_j P(h_j \mid \mathbf{v}) \tag{2.40}$$

If using Logistic units (binary case),

$$P(v_i = 1 \mid \mathbf{h}; \boldsymbol{\theta}) = s(a_i + \sum_{j=1}^{F} W_{ij} h_j) \tag{2.41}$$

$$P(h_j = 1 \mid \mathbf{v}; \boldsymbol{\theta}) = s(b_j + \sum_{i=1}^{D} v_i W_{ij}) \tag{2.42}$$

where $s(x) = 1/1 + exp(-x)$ is the logistic function.

Moreover, in some real applications (e.g. the time series analysis) with real-value observations, the RBM is generalized to Gaussian RBM (GRBM) with Gaussian visible units. The corresponding energy function is as follows:

$$E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = \sum_{i=1}^{D} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{F} b_j h_j - \sum_{i=1}^{D} \sum_{j=1}^{F} \frac{v_i w_{ij} h_j}{\sigma_i} \tag{2.43}$$

where $\mathbf{v} \in \mathbb{R}^D$ is the Gaussian visible units and $\mathbf{h} \in \{0, 1\}^F$ is a vector of binary hidden units. $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{a}, \mathbf{b}, \boldsymbol{\sigma}\}$ are the model parameters.

Like the binary case, the conditional distribution w.r.t. visible units $\mathbf{v}$ and hidden units $\mathbf{h}$ can be easily derived as follows:

$$P(v_i \mid \mathbf{h}; \boldsymbol{\theta}) = \mathcal{N}(a_i + \sigma_i \sum_{j=1}^{F} W_{ij} h_j, \sigma_i^2) \tag{2.44}$$

$$P(h_j = 1 \mid \mathbf{v}; \boldsymbol{\theta}) = s(b_j + \sum_{i=1}^{D} v_i W_{ij} / \sigma_i) \tag{2.45}$$

Parameter Estimation

The learning of parameters $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{a}, \mathbf{b}, \boldsymbol{\sigma}\}$ follows a maximum likelihood learning rule introduced in (Hinton & Sejnowski 1986). And each parameter $\theta_* \in \boldsymbol{\theta}$ can be estimated by minimizing the following negative log-likelihood:

$$-\frac{\partial logp(\mathbf{v}; \boldsymbol{\theta})}{\partial \theta_*} = \mathbb{E}_{P(\mathbf{h}|\mathbf{v})}(\frac{\partial E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})}{\partial \theta_*}) - \mathbb{E}_{P(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})}(\frac{\partial E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})}{\partial \theta_*}) \tag{2.46}$$

The first term on the right hand, a.k.a. data-dependent expectation, is tractable but the second term, a.k.a. model-dependent expectation is intractable and must be approximated (Bengio et al. 2013). In practice, as an approximation of the log-likelihood gradient, Contrastive Divergence (CD) (Hinton, Osindero & Teh 2006) has been found to be a successful tool to approximate the expectation, with a short $k - step$ (e.g. $k = 1$) Gibbs chain, denoted as CD_k.

By using CD_k, each parameter $\theta_* \in \boldsymbol{\theta}$ can be estimated as:

$$\theta_* \leftarrow \theta_* - \alpha(\frac{\partial E(\mathbf{v}^0, \mathbf{h}^0)}{\partial \theta_*} - \frac{\partial E(\mathbf{v}^k, \mathbf{h}^k)}{\partial \theta_*}) \tag{2.47}$$

where $\mathbf{v}^0$ are the visible data, $\mathbf{h}^0$ is sampled by Equation (2.42) or (2.45), and $\mathbf{v}^k$ and $\mathbf{h}^k$ are sampled from the k-step Gibbs chain.

C. Applications

Since 2006, deep belief networks have been applied and these have been successful in many areas. They have been applied not only in classification tasks such as image generation and recognition (Bengio, Lamblin, Popovici & Larochelle 2007, Ahmed, Yu, Xu, Gong & Xing 2008) and handwritten digit classification (Lee, Grosse, Ranganath & Ng 2009), but also in textures modeling (Osindero & Hinton 2008), human motion modeling (Taylor, Hinton & Roweis 2006), speech recognition (Sainath, Kingsbury, Ramabhadran, Fousek, Novak & Mohamed 2011), language processing (Mnih & Hinton 2009), robotic modeling (Hadsell, Erkan, Sermanet, Scoffier, Muller & LeCun 2008), group recommendation, and regression (Hinton & Salakhutdinov 2008).

2.4 Summary

This chapter introduces the CBA, cross-market problems and methods related to this thesis. Section 2.1 introduces behavior, coupled behavior and coupled behavior analysis. Section 2.2 presents two major applications in cross-market analysis, including financial crisis detection and financial market trend forecasting. The state-of-the-art approaches and literature in relation to the two applications are listed. Also, the limitations of existing approaches are discussed. Section 2.3 presents the building blocks of methods used in this thesis, including CHMM (Section 2.3.1), SSM (Section 2.3.2) and DBN (Section 2.3.3).

Chapter 3

Financial Crisis Detection via Coupled Market Behavior Analysis

Financial crisis detection is a widely explored and continually researched topic owing to the huge impact it can make. However, it is a challenge area of research because of the following two major reasons: the non-linear and non-stationary characteristics of financial crisis. More importantly, the instruments in different markets such as gold price and stock market index are often coupled, and a financial crisis may significantly change the couplings between different market indicators. All of these challenge most of the existing methods which rely on selecting individual indicators associated with one market, and the linear assumptions behind the model. In addition, the current methods pay no attention to the couplings across the financial markets. Thus, in this chapter, we present a new approach to detect financial crisis based on coupled market behavior analysis. This method allow us to detect financial crisis through exploring the different coupled behaviors among major global financial markets in crisis and non-crisis periods. Here coupled behaviors refer to the intra- and inter- relations between financial markets.

A financial crisis detection framework based on Coupled Hidden Markov Model is introduced to characterize the coupled market behaviors, and then to detect whether the coming period is in the crisis set or not. The framework is applied to analyze coupled behaviors among stock, commodity and interest markets as case studies. Experiments with 20 years data (from 1990 to 2009) demonstrate that the proposed approach leads to better financial crisis detection as compared to other traditionally used approaches, such as Signal, Logistic and ANN approaches.

3.1 Background and Overview

The impact of financial crisis is often disruptive on multiple perspectives, including economy, living, society and globalization. As we just experienced, the subprime mortgage crisis triggered in the US in 2007 has causes a chain of destructive effect on instruments in the US, global financial markets, and other markets and areas (Longstaff 2010). This clearly discloses the need of substantial efforts to be made on early prediction of financial crisis.

However, the effective detection of possible financial crisis is not a trivial task. Firstly, crisis has a strong transfer effect from one aspect to another, namely we can not use the change of a single indicator to represent the crisis. This is because each indicator would reflect differently to a crisis. It indicates that the simple signal approach illustrated by Kaminsky and Reinhart in (Kaminsky & Reinhart 1999) is not operable. Secondly, the financial crisis is a rare event with non-linear feature, which means that it is difficult for some approaches with linear assumption (for example, logistic approach) to capture the nature of crisis. Thirdly, financial crisis is a complex problem triggered, associated or reflected via many factors. Even with multiple indicators, it is essential to consider the coupling relationships between the indicators. Often there are dependency between indicators from various financial markets, crisis effect is passed from one market to another reflected through the couplings and indicator dynamics. It is assumed that the couplings and indicator dynamics behave significantly differently between the crisis and non-crisis periods, as the drivers triggering the crisis likely change the coupling from 'normal' to 'abnormal' status. Currently, there is limited work reported on capturing the couplings and coupling changes cross market indicators. Lastly, most of existing models rely on the selection of proper indicators. A corresponding issue is how to select discriminative indicators that are sensitive to crisis.

The above issues surrounding the existing work make it very necessary to develop new approach for effectively selecting the appropriate indicators, catering for the nonlinear dynamics, and considering their coupling relationships. For this, we propose a coupled market behavior based framework to detect financial crisis. The coupled market behaviors refer to behaviors of different market instruments, such as gold price and stock market index, which show strong coupling relationships. The framework works on the assumption that financial crisis is better reflected through coupled market behaviors rather than single indicators without coupling relationships, and the couplings change with the occurrence of financial crisis, namely sharing different coupling dynamics during and outside the crisis period. This is acceptable according to domain knowledge and the cross-market theory.

The framework consists of three stages. It firstly converts the common transnational data into a new data structure better fitting the model used in the second stage. The second stage captures the couplings between market behaviors and model the coupling difference during and outside crisis respectively through a Coupled Hidden Markov Model (CHMM). The CHMM captures the non-linear coupling relationships in multiple processes and its transitional effect from one state to another. Subsequently, the third stage detects financial crisis by observing significant difference occurring between the crisis and non-crisis periods.

The rest of this chapter is organized as follows. In Section 3.2, the coupled market behaviors are illustrated by a case study in financial markets and the corresponding problem of coupled market behavior based crisis detection is defined. The framework of our model is described in Section 3.3. Empirical outcomes and evaluation are illustrated in Section 3.4. Conclusions and future work are discussed in Section 3.5.

3.2 Problem Statement

3.2.1 A Case Study

During the 2007 global financial crisis, we can find some interesting phenomenons in different financial markets. In commodity market, for crude oil price, the 2008 calendar year was one of the most volatile periods in the history, the price reached its record high of 147 dollar per barrel in July and dropped to 60 dollar in November in the same year. Similarly, the equity market also suffered a big decline with high volatility, the S&P 500 index dropped from 1267.38 in July to 896.24 in November. In addition to that, with the fear of global recession leaded by the troubled US economy, interest market also changed in the financial crisis period by the effort of government, interest in many countries reached its record low in the period.

The above examples show that, although all three markets had great change during the financial crisis period, there exists some linkage between commodity, equity and interest markets. Such coupling is more specifically and intuitively demonstrated through Figure 3.1 and Table 3.1, and the selected indicators for these three markets are as follows.

- Commodity market: *a.The Gold price (USD per ounce, London PM fix).* It constitutes the main commodity market and often used as refuge for safe assets during financial crisis period. *b. The WTI Crude Oil Futures Price (USD per barrel).* It has become a major commodity market, not just for commodity producers, but also for investors (Daskalaki & Skiadopoulos 2011).

- Stock market: *a. The S&P 500 index.* It is a stock market index based on the market capitalizations of 500 leading companies publicly traded in the U.S. stock market. It represents the liquidity of market and recognized as one of the most commonly followed equity indices. *b. Dow Jones Industrial Average (DJIA).* It is an index that shows how 30 large publicly owned companies based in the U.S. have traded during a standard trading session in the stock market (OSullivan & Sheffrin 2007).

- Interest market: *a. The TED Spread.* It is the difference between the 3-month interest rates on interbank loans and on 3-month Treasury bill rate. It represents the counterparty risk from one bank lending to another and it is an indicator of credit risk in the economy (Brunnermeier 2008). *b. The Baa Spread.* It is the difference between the Baa Corporate bond rate and 10 year Treasury bill rate. It is wide recognized to provide an assessment of risk for investing.

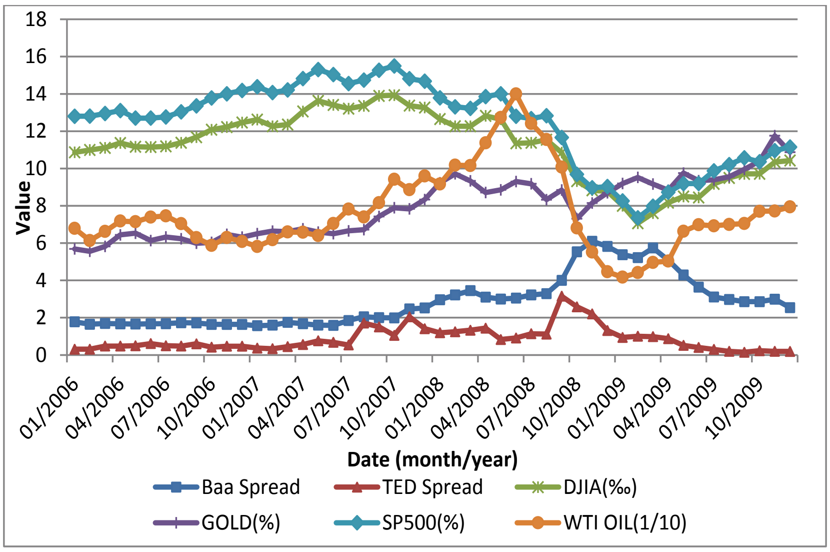


Figure 3.1: Trends of Indicators in Three Markets (2006-2009)

The data for Figure 3.1 and Table 3.1 is from January 2006 to December 2009, including a crisis period and a non-crisis period. Figure 3.1 shows

Table 3.1: Pearson Coefficients of Indicators in Three Markets (2006-2009)

		Baa Spread	TED Spread	DJIA	GOLD	SP500	WTI OIL
Baa Spread	Pearson Correlation	1	.987**	-.060	.282**	-.129**	-.250**
	Sig.(2-tailed)		.000	.080	.000	.000	.000
TED Spread	Pearson Correlation	.987**	1	-.078*	0.258**	-.133**	-.255**
	Sig.(2-tailed)	.000		.024	.000	.000	.000
DJIA	Pearson Correlation	-.060	-.078*	1	0.234**	-.016	-.018
	Sig.(2-tailed)	.080	.024		.000	.647	.595
GOLD	Pearson Correlation	.282**	0.258**	0.234**	1	-.040	-.068*
	Sig.(2-tailed)	.000	.000	.000		.242	.049
SP500	Pearson Correlation	-.129**	-.133*	-.016	-.040	1	.258**
	Sig.(2-tailed)	.000	.000	.647	.242		.000
WTI OIL	Pearson Correlation	-.250**	-.255**	-.018	-.068*	258**	1
	Sig.(2-tailed)	.000	.000	.595	.049	.000	

** Correlation is significant at the 0.01 level(2-tailed). * Correlation is significant at the 0.05 level(2-tailed).

that the relationships between these indicators fluctuated during the two periods. The relationships remain much more stable in the non-crisis period (before the late 2007) than the crisis period. Table 3.1 further demonstrates that there are strong correlations between the indicators through the whole period, as shown in the Pearson correlations. Based on this, we can come to the conclusion that these three markets are coupled with each other, but the coupled relationships behave differently between the crisis period and non-crisis period.

This typical case study supports our assumption that financial crisis has the transfer effect on multiple indicators, which display different nonlinear and dynamic characteristics. They are coupled in some way, the coupling relationships change with the occurrence of crisis. To effectively detect financial crisis, it is essential to consider the the nonlinear and dynamic factors, and the couplings between indicators. A significant change of the coupling relationships can serve as a strong sign for differentiating the crisis and normal periods. Below, we take these observations to define the problem of coupled market behavior analysis for financial crisis detection, by considering the major difference on multiple indicator coupling relationships between crisis and non-crisis periods.

3.2.2 Problem Formalization

The problem of coupled behavior based financial crisis detection can be formalized as follows. Let function $f(\cdot)$ capture the coupling relationships between the three markets in the above case, $g(\cdot)$ be the corresponding objective function determining whether a crisis exists. Two models are built separately for the crisis and non-crisis periods: Model CM with the coupling function $f_{CM}(\cdot)$ characterizes the coupled market behaviors during the crisis, while Model NM with the coupling $f_{NM}(\cdot)$ describes the characteristics and coupling relationships between indicators from the non-crisis period. If

$$g_{CM}(t)(f_{CM}(\cdot)) \geq g_{NM}(t)(f_{NM}(\cdot)), \tag{3.1}$$

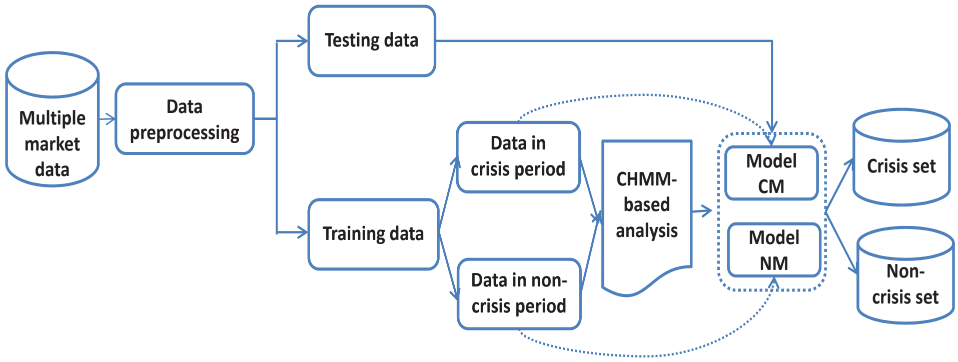


Figure 3.2: The Proposed Crisis Detection Framework

time t is in the crisis set, and otherwise in the non-crisis set.

Correspondingly, our key task in financial crisis detection is to determine the coupling function $f(\cdot)$ and the objective function $g(\cdot)$ corresponding to two different periods. Below, the Coupled Hidden Markov Model (CHMM) is explored to capture the coupling relationships and nonlinear dynamics of multiple indicators, with an objective function built to check the major difference in the CHMM outputs.

3.3 Modeling Framework

This section introduces the system framework for financial crisis detection, the data structure and conversion from business data to behavioral data, the modeling process and the detection algorithm respectively.

3.3.1 Working System

Based on the case study and corresponding problem definition in Section 3.2, we propose a CHMM-based financial crisis detection framework, which is depicted in Figure 3.2. It consists of three major stages: 1) Data preprocessing, which converts the transactional data into the behavior-oriented data structure that better fits the CHMM analysis. The transformed data is further partitioned with the involvement of domain knowledge into training and testing sets. 2) CHMM-based coupled market behavior modeling, namely using CHMM to model the coupled market behaviors in the two different periods. We explain how the nonlinear characteristics of the indicators and the couplings between indicators are captured in the CHMM models. 3) Crisis detection, which determines whether a crisis appears using the data-driven method.

3.3.2 Data Preprocessing

To better fit the model, in this stage we mainly focus on three parts: indicator selection, data normalization, and data partition.

1. Indicator selection. we use one Markov chain to represent one financial market, so we need to find one indicator for each market. In the real world, there are more than one indicator that can represent the market. For example, in Section 3.2, each of the three markets owns two major indicators. Here we select one indicator for each market which has higher correlations with other markets. This is because our focus is on the coupling relationships among various markets, the indicator more relevant with other markets encloses strong discriminative power.

Definition 3.1 *Suppose there are I markets, each market own N indicators. Market indicator correlation $MIC_{i_1j_1}$ refers to the correlations of indicator $MI_{i_1j_1}$ with indicators in other markets, where $(i_1 \neq i_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq N)$. The $corr(\cdot)$ is the pearson correlation coefficients of the two indicators.*

$$MIC_{i_1j_1} = \sum_{i_2=1}^{I} \sum_{j_2=1}^{N} corr(MI_{i_1j_1}, MI_{i_2j_2}) \tag{3.2}$$

2. Data normalization. The data we choose includes the closing prices of indicators in each market. The data types are different among the various markets, thus we normalize the original price series into $[0, 1]$ to better fit

the model. The price of indicator at time t PI_t is normalized to PI'_t by

$$PI'_t = (PI_t - PI_{min})/(PI_{max} - PI_{min}) \tag{3.3}$$

where PI_{min} and PI_{max} are the minimum and maximum prices, respectively.

3. Data partition. In our system, the training data would be divided into two parts: crisis set and non-crisis set. Two models are trained on them respectively. Model CM represents the complex coupled market behaviors in the crisis period, Model NM indicates the relationships in the non-crisis stage. Domain knowledge from the National Bureau of Economic Research (NBER) Business Cycle Dating Committee[1] is involved in this data splitting. The following economic cycles occurred in our selected data: Q3-1990 to Q1-1991 (led by the Gulf war), Q1-2001 to Q4-2001 (triggered by the dot-com bubble) and Q4-2007 to Q2-2009 (caused by the subprime crisis). The former two cycles are used for training, with the last for testing.

3.3.3 CHMM-based Market Behavior Modeling

In the CHMM-based financial crisis detection system, three HMM chains, namely HMM-C capturing the commodity market sequence $\Phi(C)$, HMM-S enclosing the stock market sequence $\Phi(S)$, and HMM-I for the interest market sequence $\Phi(I)$, are built. Accordingly, a CHMM is built to incorporate all three sequences, which maps the financial crisis problem represented by coupled market behaviors of three markets into the CHMM. This mapping

[1] The NBER's Business Cycle Dating Committee maintains a chronology of the U.S. business cycle. Available at http://www.nber.org/cycles.html

works as follows:

$$Financial\ crisis\ analysis \rightarrow CHMM\ modeling \tag{3.4}$$

$$f(\cdot) \rightarrow \Omega(A, B, R, \pi) \tag{3.5}$$

$$\Phi(\cdot)|observation \rightarrow A \tag{3.6}$$

$$\Phi(\cdot)|transition \rightarrow B \tag{3.7}$$

$$r_{ij} \rightarrow R \tag{3.8}$$

$$\Phi(\cdot)|prior \rightarrow \pi \tag{3.9}$$

where $i, j \in \{S, C, I\}$, r_{ij} represents the coupling coefficient of two markets i and j.

Two CHMM models are trained on the Q3-1990 to Q1-1991 and Q1-2001 to Q4-2001 data sets respectively. Below, we discuss the calculation of the parameters in the CHMM.

Assume there are C coupled chains, the hidden state is denoted by Z, the state transition probability is[2]

$$P(Z_t^{(c)}|Z_{t-1}^{(1)}, Z_{t-1}^{(2)}, \ldots, Z_{t-1}^{(C)}) \tag{3.10}$$

where the $Z_t^{(c)}$ is the hidden state of model c at time t. We can see from above equation that the number of parameters is H^C when the number of hidden state is H for each chain. This is not easy to do parameter learning. Based on this, many researchers proposed several variations of CHMM wherein the inference problems are more tractable. Authors in (Cao et al. 2012) used the method proposed by (Brand 1997). According to this method, the state transition probability become the product of all marginal conditional probabilities,

$$P(Z_t^{(c)}|Z_{t-1}^{(1)}, Z_{t-1}^{(2)}, \ldots, Z_{t-1}^{(C)}) = \prod_{c'} P(Z_t^{(c)}|Z_{t-1}^{(c')}) \tag{3.11}$$

The above method reflects an approximation, which is not a properly defined probability density, namely the right hand side does not sum up to one (Zhong & Ghosh 2002).

[2]This chapter focus on the type of CHMM that the state of one chains at time t depends on the states of all chains (including itself).

In this chapter, we use a method illustrated in (Zhong & Ghosh 2001), the authors introduce new parameters to capture the interaction and model the joint transition probability as

$$P(Z_t^{(c)}|Z_{t-1}^{(1)}, Z_{t-1}^{(2)}, \ldots, Z_{t-1}^{(C)}) = \sum_{c'=1}^{C}(r_{c'c}P(Z_t^{(c)}|Z_{t-1}^{(c')})) \tag{3.12}$$

where $r_{c'c}$ is the coupling coefficient which can measure the coupling weights from chain c' to c, namely how much $Z_{t-1}^{(c')}$ affect $Z_t^{(c)}$. The authors model the joint dependency as a linear combination of all marginal dependencies. An simple interpretation is as follows:

$$\begin{aligned} P(y|x_1, x_2, \ldots, x_C) &= \frac{P(y, x_1, x_2, \ldots, x_C)}{P(x_1, x_2, \ldots, x_C)} \\ &= \frac{P(x_1), P(y|x_1), \ldots, P(x_C|P(y, x_1, x_2, \ldots, x_C))}{P(x_1, x_2, \ldots, x_C)} \\ &= \omega_1 P(y|x_1) \end{aligned} \tag{3.13}$$

where y is the represent the current state and x is for previous states simplicity, $\omega_1 = \frac{P(x_1),\ldots,P(x_C|P(y,x_1,x_2,\ldots,x_C))}{P(x_1,x_2,\ldots,x_C)}$. Similarly, we can obtain

$$\begin{aligned} P(y|x_1, x_2, \ldots, x_C) &= \omega_1 P(y|x_1) = \omega_2 P(y|x_2) \\ &= \cdots = \omega_C P(y|x_C) \end{aligned} \tag{3.14}$$

Then the joint conditional probability can be rewritten as

$$P(y|x_1, x_2, \ldots, x_C) = \sum_{c=1}^{C} r_c P(y|x_C) \tag{3.15}$$

where $r_c = \frac{1}{C}\omega_c, 1 \leq c \leq C$ are the parameters used to represent the interactions among different markets in our research. Then corresponding learning algorithm can be used to learning the model, more details refer to (Zhong & Ghosh 2002).

3.3.4 Financial Crisis Detection

In the above, we have got two models: models CM and NM capturing the coupled market behaviors for the crisis and non-crisis stages respectively.

Here these models are used to determine whether the coming coupled market behavior b^k appears to be in crisis or not. To this end, the conditional likelihood (CL) is calculated for the two models CM and NM corresponding to the given coupled market behavior b^k. The algorithm for detecting financial crisis is described in Algorithm 3.1. Steps 1 to 2 train the crisis model CM and non-crisis model NM. Steps 3 to 11 form a loop process to compute the likelihood of new coupled market behavior b^k in the testing set based on the two models. The output of the algorithm includes two sets: crisis set CS and non-crisis set NS.

Algorithm 3.1 Financial Crisis Detection via CHMM

Require: A training set T_1 of financial crisis data $\{b^1, b^2, \ldots, b^M\}$;A training set T_2 of non-financial crisis data $\{b^1, b^2, \ldots, b^N\}$;A testing set $\{b^1, b^2, \ldots, b^O\}$

1: Train one model CM on the training set T_1;
2: Train one model NM on the training set T_2;
3: **for** b^k in the Testing set **do**
4:     Compute the likelihood of b^k given the model CM and NM, desperately:
5:     $CL(b^k \mid CM)$ and $L(b^k \mid NM)$;
6:     **if** $L(b^k \mid CM) \geq L(b^k \mid NM)$ **then**
7:         $b^k \rightarrow CS$
8:     **else**
9:         $b^k \rightarrow NS$
10:     **end if**
11: **end for**
12: **return** A financial crisis set CS; A non-financial crisis set NS

3.4 Experiments

In this section, we discuss the data sets, baseline methods, experimental settings as well as evaluation metrics and experimental outcomes.

3.4.1 The Data Sets

The data from three financial markets: commodity market, stock market and interest market are extracted for the experiments. As shown in Table 3.2, two typical indicators are chosen for each market. According to the discussion in the data preprocessing stage in Section 3.3.2, only one indicator is selected to represent each market based on their correlations with other markets. The selected indicators are: the WTI Crude Oil Price, DJIA and the TED Spread.

The data is obtained from the Economic Research[3]. The data includes weekly prices from January 1990 to December 2009, and is divided into two parts considering the domain knowledge: the training set consists of the data from 1990 to 2005, the testing set from 2006 to 2009. By considering the financial events and performance at that time period, the time period associated with the crisis is labeled for the two parts. Models are trained on the training set to capture the characteristics of coupled market behaviors in crisis and non-crisis periods, separately. The trained models are deployed on the testing set to detect the financial crisis. The coupled market behaviors in the testing set can be seen as $b^k(1 \leq k \leq O)$ in Algorithm 3.1. Because indicators in different markets may appear on different trading days, we delete those days that some markets are missing and only choose the trading days that all markets have trading. Then the size of training set is 835 observations, and 210 observations in testing set.

3.4.2 Comparative Methods

In order to evaluate the performance of our approach in analyzing financial crisis, we compare it with the following approaches:

[3]http://research.stlouisfed.org/

Table 3.2: Trading Indicators in Three Markets

Market	Indicator 1	Indicator 2
Commodity Market	Gold price	WTI Crude Oil Price
Equity Market	S&P 500	DJIA
Interest Market	The TED Spread	The Baa Spread

- *Signal approach.* This approach (Kaminsky 1999) believes that the indicators behave differently at the moment when the financial crisis occurs when compared with a relatively normal period. Suppose there are n possible indicators, the change of an indicator $X^j(1 \leq j \leq n)$ is said to signal a financial crisis at time t when it crosses an 'optimal threshold' $\overline{X}^j$,

$$\{S_t^j = 1\} = \{S_t^j, |X_t^j| > |\overline{X}_t^j|\} \tag{3.16}$$

$$\{S_t^j = 0\} = \{S_t^j, |X_t^j| < |\overline{X}_t^j|\} \tag{3.17}$$

here $S_t^j = 1$ represents indicator j used in signaling the crisis at time t, and $S_t^j = 0$ otherwise. Usually the threshold value can be located between the 10th percentile and 20th percentile (Kaminsky 1999). However, different indicators often produce different prediction outcomes for financial crisis. To obtain a more stable prediction outcome, multiple indicators are used and combined for the signaling (Lin et al. 2008),

$$I_t = \sum_{j=1}^{n} S_t^j \tag{3.18}$$

I_t is then recognized as an indicator that combines all the information. The threshold for this indicator is determined in the same way as in the case of other indicators.

- *Logistic approach.* The Logistic regression is used for predicting the outcome of a categorical dependent variable based on one or more predictor variables. Suppose $Y_t = 1$ represents that there is a financial crisis at time t, and $Y_t = 0$ otherwise. P_t is the probability of having a financial crisis at time t,

$$\begin{aligned} P_t &= P(Y_t = 1) = E(Y_t|X) \\ &= \frac{1}{1 + e^{-b_0 + b_1 x_1 + \cdots + b_n x_n + \varepsilon}} \end{aligned} \tag{3.19}$$

here $x_i (1 \leq i \leq n)$ is explanatory variable, ε is the error term. Then the log-likelihood function can be written as follows:

$$l(\theta) = \sum_{t=1}^{T} Y_t(ln(P_t)) + (1 - Y_t)(ln(1 - P_t)) \tag{3.20}$$

where T is the number of periods. The parameters can be obtained through estimating the maximum likelihood.

- *ANN approach.* An ANN network consists of an interconnected group of artificial neurons and processes information using a connectionist approach, with functions to map input values into output values. Using ANN to analyze financial crisis is the process of learning to separate the testing set into different classes (crisis set and non-crisis set) by finding common features between samples in the training set. Here we will use the most popular ANN learning algorithm, the back-propagation algorithm (Kim, Oh, Sohn & Hwang 2004), to conduct the learning process.

3.4.3 Experimental Settings

All the comparison approaches use the same indicators as our proposed CHM-M approach. Also, the training data and testing data are the same for the Logistic approach, ANN and CHMM. Since the signal approach only considers the simple calculation of the number of indicators in signaling, no relationships among the indicators are considered. The Logistic approach only captures the linear relationships between indicators, while ANN captures the non-linear relationships without considering the coupling relationships between indicators.

3.4.4 Evaluation Metrics

In this chapter, to compare the performance of our approach and other approaches, we use following evaluation methods:

- *Overall Accuracy.* Accuracy is the percentage of correctly classified instances.

$$Overall\ Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \tag{3.21}$$

  where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively, we treat financial crisis cases as the positive class here.

- *Precision.* Precision is the percentage of correctly classified positive instances.

$$Precision = \frac{TP}{TP + FP} \tag{3.22}$$

- *Type I error.* The percentage of the number of times with no signal fired when there is a crisis against the times that there is a crisis.

- *AUC.* AUC is the area under the ROC curve, where ROC is created by plotting the TP rate (true positive out of the positives) and FP rate (false positive out of the negatives). The AUC represents the classification accuracy, the larger the better.

3.4.5 Experimental Results

A. Technical Performance

Here we compare the technical performance of our approach and other three approaches on the testing data. Overall accuracy, precision and type I error listed in former part are calculated and the results are reported in Table 3.3, Figure 3.3 and Figure 3.4. The horizontal axis (P-Num) in Figure 3.3 and Figure 3.4 stands for the number of detected financial crisis (i.e., the number of trading weeks with abnormal coupled market behaviors).

Table 3.3: Technical Performance of Four Approaches for Financial Crisis Detection

Approach	Overall Accuracy	Precision	Type I error
Signal	0.5714	0.6038	0.3962
Logistic	0.7895	0.6585	0.3415
ANN	0.6857	0.7170	0.2830
CHMM	**0.8571**	**0.9057**	**0.0943**

The results in Table 3.3, Figure 3.3 and Figure 3.4 come up with the following conclusions: Our CHMM-based financial crisis detection approach performs best in terms of all data sets and evaluation metrics. For instance, CHMM outperforms other methods by greatly reducing the Type I error to about 18.9% to 30.2% of the other methods. CHMM is with the highest accuracy increase about 36% compared to the Signal method, and 20% over the best performing Logistic method when P-Num = 45. Similarly, the precision improvement could be as high as about 40% against the Signal method, and 20% against the best performing Logistic when P-Num = 45. All these clearly show that the coupled market behavior analysis is a promising approach for financial crisis detection. The main reason lies on that our approach can capture the complex coupling relationships among financial markets in financial crisis and non-crisis periods. In addition, CHMM has been demonstrated to be an useful model to characterize the coupled market behaviors.

Interestingly, the results of the logistic and ANN approaches are conflicting with each other. In the whole testing period, the logistic approach performs better than ANN by accuracy, but the ANN gets a better precision. This may be because, in real financial markets, the relationships among different markets may disclose more stationary and linear characteristics in the non-crisis period, while such a feature becomes more non-linear and dynamics when a crisis takes place. However, both of them overlooks the couplings between market indicators and the coupling changes during crisis. The Signal approach achieves the worst performance, indicating that we cannot detect

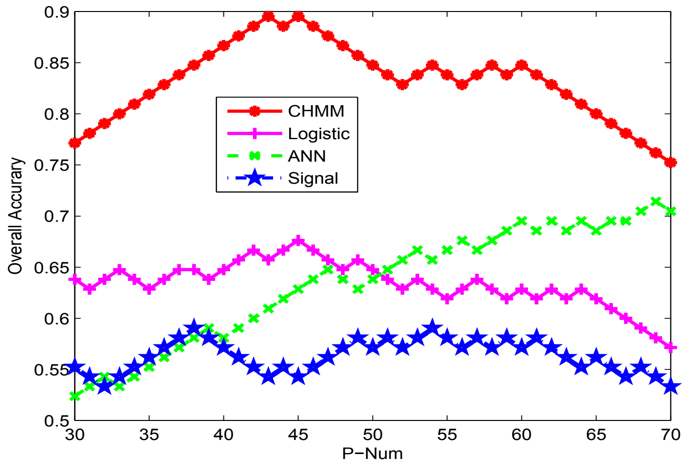


Figure 3.3: Accuracy of Four Approaches for Financial Crisis Detection

financial crisis only through observing simple changes of indicators.

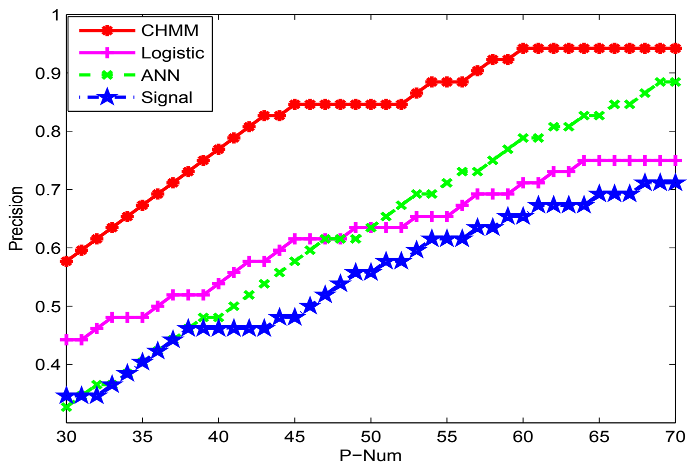


Figure 3.4: Precision of Four Approaches for Financial Crisis Detection

B. Exploration of Performance Score

In this part we compare the four approaches in terms of three performance scores, denoted as '1-AUC' '2-Accuracy' and '3-Precision'. The results are depicted in Figure 3.5. The results show that the CHMM-based method outperforms all the rest on all performance aspects. This further shows the great potential of incorporating couplings between market behaviors in detecting financial crisis.

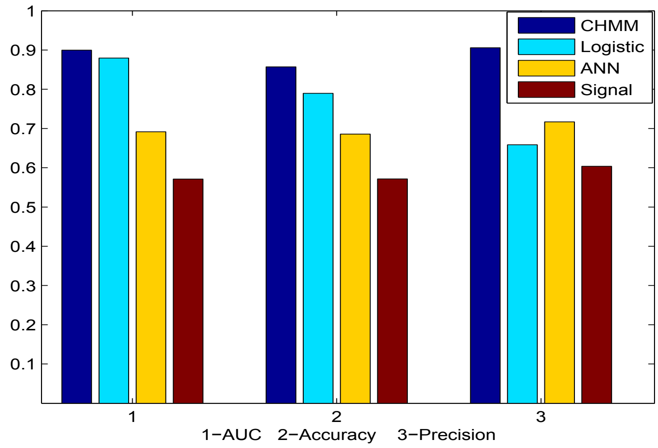


Figure 3.5: Performance Scores of Four Approaches for Financial Crisis Detection

3.5 Summary

The effective detection of financial crisis is crucial but very challenging. Many methods have been studied by applying financial and statistical theories. A modern trend is the data-driven learning approach, which incorporates the advanced learning techniques to identify inconsistency caused by crisis on the major market indicators. In this chapter, we propose a new financial crisis detection approach to consider the nonlinear characteristics in the market, the coupling relationships between different market behaviors, and the significant impact caused by crisis occurrence on the indicator dynamics and couplings. A CHMM-based model is designed to capture the above aspects and deployed to detect the subprime mortgage crisis by selecting major indicators from the commodity, stock and interest markets. The results show the clear advantage of our approach against the Signal, Logistic and ANN based methods with a significant accuracy improvement.

Chapter 4

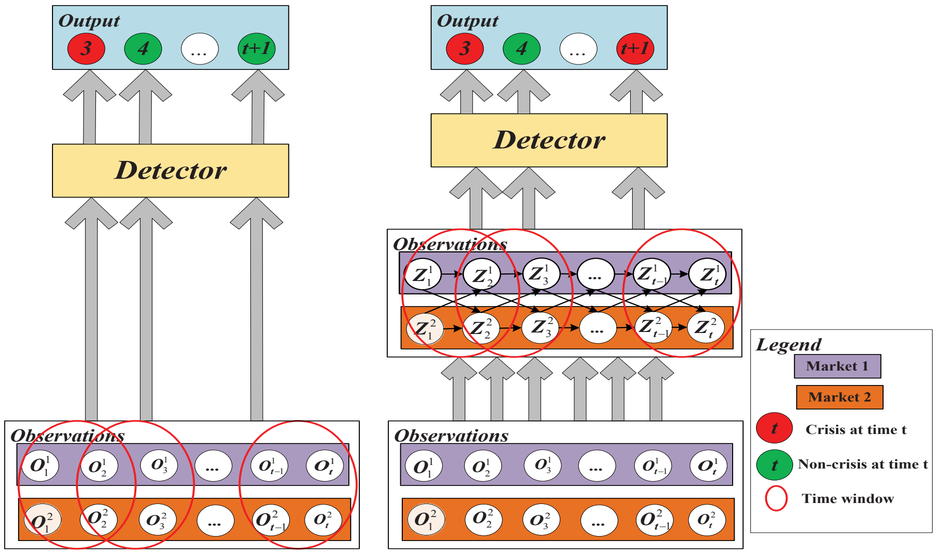
Financial Crisis Forecasting via Coupled Market State Analysis

It is well known that financial crisis forecasting is not a trivial task. The current approaches, such as the Logistic-based detector and ANN-based detector, conduct the forecasting directly from data/observations. This may lead to biased results since financial crisis is a nonlinear and dynamic issue and it is often driven by particular couplings behind the observation. Once the couplings change, the observations fluctuate.

Based on this, in this chapter we investigate and forecast financial crisis through teasing out the coupled relationships between hidden market states. Here, coupled market states refer to a set of dynamic hidden states in different types of global financial markets, which are used to represent the hidden state transitions generated by interactions between different markets. Accordingly, a coupled state space model is introduced to capture the coupled market states, and the coupled market states are fed into a traditional detector as features. Substantial experiments on 20 years of real financial market data verify our assumption. The experimental results show the clear advantage of our proposed approach against the traditional observation-based approaches.

4.1 Background and Overview

To date, the increasing number of researchers and practitioners have recognized the need and challenge of financial crisis forecasting. However, most current approaches directly conduct the prediction through observations of indicators. As shown in Figure 4.1 (a), a crisis detector (e.g. a Logistic regression classifier), is used to forecast crisis for time $t+1$ based on observations of all market indicators from time $t-k+1$ to t (k denotes the time window size). We argue that such kind of approaches overlook the underlying coupled relationships between markets, which cannot be directly detected. Once the couplings changed, the observations will be fluctuated.



(a) Observation Based Forecasting (b) CMSA Based Forecasting

Figure 4.1: Two Types of Financial Crisis Forecasting

To represent the hidden couplings between various markets, we propose a new forecasting framework based on *Coupled Market State Analysis (CMSA)* to predict the financial crisis. Here the *coupled market state (CMS)* refer to a set of dynamic hidden states from various markets, which are used to represent the state transition induced by the constant interactions between markets. Our proposal works on the assumption that market indicators are governed by a collections of *CMSs* which are the better features to capture financial crisis.

Accordingly, we insert a *Coupled State Space Model (CSSM)* between observations and detector layers, as illustrated in Figure 4.1 (b), so as to conduct the *CMSA* over all markets. As a whole, the working process of our approach can be described as follows: firstly, we learn the *CMSs* behind the observations using the *CSSM*; secondly, the *CMSs* are fed into a detector (same as Figure 4.1 (a)) as features; finally, the detector performs forecasting to output the results of whether in financial crisis set. Compared with the observation-based forecasting approach, we can see that the major difference reflects the the effect of *CMSA*.

The chapter is organized as follows. The corresponding problem is formalized in Section 4.2. Section 4.3 proposes the framework of the *CMSA*, which includes the definitions of *CMSs* and *CMSA*, the exploration of *CSS-M*, the modeling process and the forecasting process. Section 4.4 shows the experimental outcomes and evaluations. Finally, we conclude this work and address future work in Section 4.5.

4.2 Problem Formalization

This chapter aims to effectively forecast financial crisis through considering the change of the couplings between the hidden states in various markets. A significant change of the coupled market states can serve as a strong sign for differentiating the crisis and normal periods. Then the forecasting problem can be formalized as follows: $f(\cdot)$ is a function used to capture the complex coupled relationships between market states in different financial markets, and an objective function $g(\cdot)$ is built to forecast the possibilities of crisis and non-crisis. If at time t,

$$g_{t+1}^{D}(c=1) \mid f_{t-k+1}^{t}(CMS) \geq g_{t+1}^{D}(c=0) \mid f_{t-k+1}^{t}(CMS), \tag{4.1}$$

then time $t+1$ is a crisis period, otherwise a non-crisis period. Here $g_{t+1}^{D}(c=1) \mid f_{t-k+1}^{t}(CMS)$ represents the possibility of a crisis by using a detector with the *CMSs* from time $t-k+1$ to time t (k denotes the window size).

In order to do the forecasting, our key task now is to build a proper model to determine the specific function $f(\cdot)$ and the corresponding objective function $g(\cdot)$. Below, *CMSA* and *CSSM* are introduced to understand the *CMS*, and then CHMM is explored as a specific *CSSM* to capture the coupled market states relations.

4.3 Modeling Framework

In order to resolve the above problem, we propose a *CMSA* based financial crisis forecasting framework, which is depicted in Figure 4.2. It consists of three major steps and we discuss them in detail hereafter.

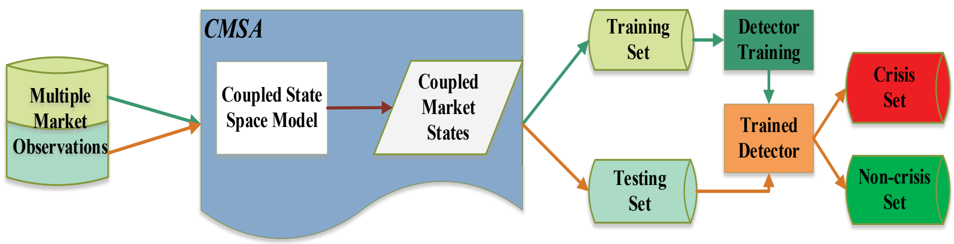


Figure 4.2: *CSSM*-based Financial Crisis Forecasting Framework

4.3.1 Coupled Market State Analysis

A. Definitions

Coupled market state (CMS) refer to the hidden states from multiple markets with inter- and intra- relations. Suppose there are I markets $\{M_1, M_2, \ldots, M_I\}$, a market M_i undertakes J market states $\{MS_{i1}, MS_{i2}, \ldots, MS_{iJ}\}$. Then a Market State Behavior Feature Matrix $FM(MS)$ is defined as follows:

$$FM(MS) = \begin{pmatrix} MS_{11} & MS_{12} & \dots & MS_{1J} \\ MS_{21} & MS_{22} & \dots & MS_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ MS_{I1} & MS_{I2} & \dots & MS_{IJ} \end{pmatrix}$$

Then the intra-couplings within each market sate is the relationships within one row of the above matrix, while how the states interact between different markets are embodied through the columns of $FM(MS)$, indicated as inter-couplings.

Definition 4.1 *Coupled Market State* CMS *refer to market states* $MS_{i_1j_1}$ *and* $MS_{i_2j_2}$ *that are coupled in terms of intra-coupling and inter-coupling.*

$$\theta(MS) = \{MS_{i_1j_1} \odot MS_{i_2j_2} | i_1 = i_2, 1 \leq i_1, i_2 \leq I\} \tag{4.2}$$

$$\eta(MS) = \{MS_{i_1j_1} \odot MS_{i_2j_2} | i_1 \neq i_2, 1 \leq i_1, i_2 \leq I\} \tag{4.3}$$

where $\theta(\cdot)$ *is the intra-coupling function and* $\eta(\cdot)$ *represents the inter-coupling function.* $\odot$ *means the interactions of* $MS_{i_1j_1}$ *and* $MS_{i_2j_2}$*.*

Definition 4.2 *Coupled market state analysis* CMSA *is to build the objective function* $g(\cdot)$ *under the condition that the hidden market states are coupled with each other by coupling function* $f(\cdot)$*, and satisfy the following conditions:*

$$f(\cdot) ::= \{\theta(MS), \eta(MS)\} \tag{4.4}$$

$$argmax_c g(c) | f(\cdot) \tag{4.5}$$

where $c \in \{0, 1\}$ *in this chapter,* 0 *represents the non-crisis set, while* 1 *denotes the crisis set.*

The above discussion give the basic concept of *CMSA*, and the representations show us the road map from understanding to modeling *CMSA* in the real financial market. In the following section, we discuss the specific mapping and modeling issues.

B. Coupled State Space Model

Definition 4.3 *Coupled state space model* CSSM *refers to a type of graphic model not only describes the latent variable and the observed measurement, but also capture the coupled relationships between latent variables across different state space models (SSMs)*[1]*. Suppose there are C SSM, the basic state space equations are:*

$$Y_t^c = H_t U_t^c + v_t \tag{4.6}$$

$$U_t^c = F(\theta(U_{t-1}^c), \eta(U_{t-1}^{c'})) \tag{4.7}$$

then $U_t = [U_t^1, U_t^2, \cdots, U_t^C]$. Here $(1 \leq c', c \leq C) \wedge (c \neq c')$, F is a transition function, $\theta(\cdot)$ is build to capture the intra relations in a same SSM, while $\eta(\cdot)$ encodes the couplings between different SSMs.

In this chapter we use CHMM as a concrete implementation of *CSSM* to learn coupled states between markets, since CHMM is a probabilistic state space model to capture the non-linear coupling relationship in multiple processes and its transitional effect from one hidden state to another (Zhong & Ghosh 2001).

4.3.2 Modeling Process

A. Indicator Selection

In CHMM, we use one Markov chain to represent one financial market, so here we select one indicator for each market which has higher correlations with other markets by following equation:

$$MIC_{i_1 j_1} = \sum_{i_2=1}^{I} \sum_{j_2=1}^{N} corr(MI_{i_1 j_1}, MI_{i_2 j_2}) \tag{4.8}$$

Indicator correlation $MIC_{i_1 j_1}$ refers to the correlations of indicator $MI_{i_1 j_1}$ with indicators in other markets $MI_{i_2 j_2}$, where I is the number of markets,

[1]Details of SSM refer to Section 2.3.2

and each market owns N indicators, $(i_1 \neq i_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq N)$. Here $corr(\cdot)$ is the Pearson correlation coefficient of the two indicators.

B. CHMM-based CMSA Modeling

There are two mapping processes in this section: one from CMSA to CSSM, namely from formalized issue to an abstracting model; another from CSSM to CHMM, which resolve the issue with a specific tool CHMM.

As mentioned above, there are three market state sequences: $\Phi(S)$ enclosing the stock market state sequence, $\Phi(C)$ and $\Phi(I)$ representing commodity and interest market state sequences, separately. $\{Z_1, Z_2, \ldots, Z_H\}$ is a set of hidden states, where z_t is the hidden state at time t. $\{X_1, X_2, \ldots, X_V\}$ is a set of observation symbols, $O = \{O_1, O_2, \ldots, O_T\}$ is an observation sequence, o_t is the observation at time t. The specific mapping relations are illustrated in Figure 4.3.

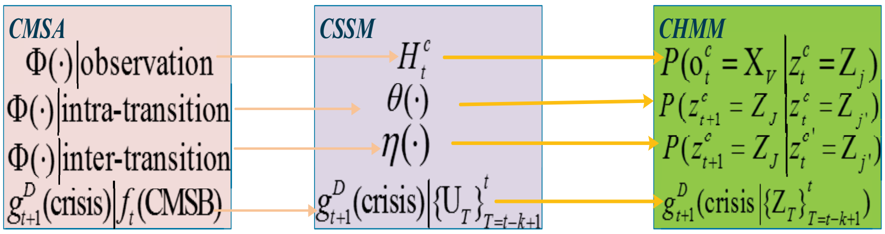


Figure 4.3: Two Mapping Processes

4.3.3 Forecasting Process

Figure 4.4 illustrates the general framework of the proposed forecasting process. For each observation interval $O_{t-k+1:t}$ (k is the time window), the first step is to train the CHMM using the k observations ($O_{t-k+1:t}$) in the three market to obtain corresponding market states $Z_{t-k+1:t}$. Then based on the coupled states, the trained detector give the probabilities of crisis $(P^D_{t+1}(c = 1|\{Z_T\}^t_{T=t-k+1}))$ and non-crisis $(P^D_{t+1}(c = 0|\{Z_T\}^t_{T=t-k+1}))$. After compared the two probabilities, we can find whether time $t+1$ is in a financial crisis set.

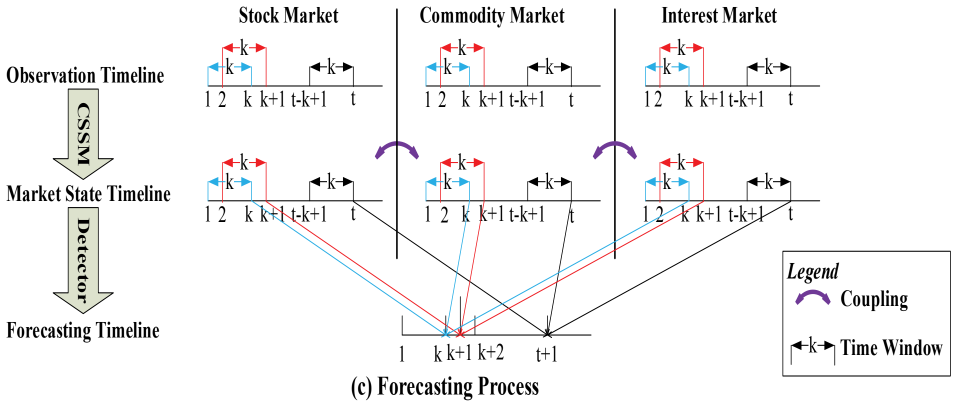


Figure 4.4: *CSSM*-based Financial Crisis Forecasting Process

4.4 Evaluation and Discussion

4.4.1 The Data Sets

The data from three financial markets: commodity market, stock market and interest market are extracted for the experiments. We select one indicator for each market according to Equation (4.8). The selected indicators are: DJIA, the WTI Crude Oil Price and the BAA Spread.

The data set used in this section includes weekly closing prices from Jan 1990 to Dec 2010 (1096 observations), obtained from the Economic Research (http://research.stlouisfed.org/), and the prices are decoded into return as the symbols which can be calculated by $RI_t = \frac{PI_t - PI_{t-1}}{PI_{t-1}} * 100\%$, here RI_t and PI_t are, respectively, the return and closing price at time t. Here we divide the data into two parts: training set from Jan 1990 to Dec 2006 (887 observations), testing set from Jan 2007 to Dec 2010 (209 observations). According to the domain knowledge from the National Bureau of Economic Research (N-BER) Business Cycle Dating Committee (http://www.nber.org/cycles.html), there are two crisis periods in training data set: July 1990 to March 1991 (led by the Gulf war), and March 2001 to Nov 2001 (triggered by the dot-com bubble), one crisis period in testing data set: Dec 2007 to June 2009 (caused by the subprime crisis). Because indicators in different markets may appear on different trading days, we delete those days on which some market data is missing and only choose the days with trading data from all financial markets.

4.4.2 Comparative Methods

- *Signal.* This basic idea of this method is the variables behaves differently in financial crisis period when compared with a normal period (Peng & Bajona 2008). We use it as a baseline method.

- *Logistic.* We use this approach with indicators from the three different markets, and the parameters can be obtained through MLE.

- *ANN.* we use the back-propagation algorithm in (Ozkan-Gunay & Ozkan 2007) with indicators from the three different markets to train the model.

4.4.3 Evaluation Metrics

The technical performance evaluation is based on following metrics:

- $Accuracy = \frac{TN+TP}{TP+FP+FN+TN}$.

- $Precision = \frac{TP}{TP+FP}$.

- $Recall = \frac{TP}{TP+FN}$.

where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively. We treat financial crisis cases as the positive class here.

4.4.4 Experimental Results

Here we compare the technical performance of our approach against the comparative approaches on the testing data with different window size k. Accuracy, precision and recall listed in the former part are calculated. The results are reported in Table 4.1 and Figure 4.5.

Table 4.1: Accuracy of Five Approaches for Financial Crisis Forecasting

Approach	Accuracy			
	$k = 1$	$k = 2$	$k = 3$	$k = 4$
Signal	0.5604			
Logistic	0.6754	0.6926	0.6953	0.7137
ANN	0.7128	0.7471	0.7617	0.7294
CHMM-Logistic	**0.7923**	**0.8132**	**0.8320**	0.7647
CHMM-ANN	0.7856	0.8016	0.8281	**0.8157**

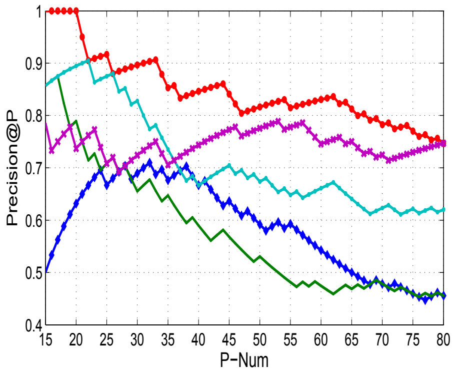
Table 4.1 shows the accuracy performance of the five approaches over the whole testing period. We can find that the baseline method of Signal does not achieve a good performance, this is because Signal is rely on selection of indicators, and pays no attention on hidden couplings between indicators. For the similar reason, the Logistic and ANN approaches do not perform very well. Note that the ANN outperforms the Logistic approach, the main reason here is the Logistic approach is under a linear assumption, but the financial crisis reveals non-linear characteristics. Our CHMM based approaches have better performance than the comparative approaches with all window sizes. For instance, the CHMM-Logistic has about 14% improvement over the Logistic approach when time window equals to 3, and CHMM-ANN has around 9% gain over the ANN method when time window equals to 4. The main reason can be interpreted as follows: unlike those methods predict financial crisis directly from data, our approach design a framework of *CMSA* to learn the hidden *CMSs* over different markets which removes the vulnerabilities of data; in addition, CHMM has been demonstrated to be a useful model

to characterize the *CMSs*. Moreover, while without inserting previous hidden states ($k = 1$), the CHMM based models still perform better than the observation based models. All these illustrate the great importance of *CMSs*.

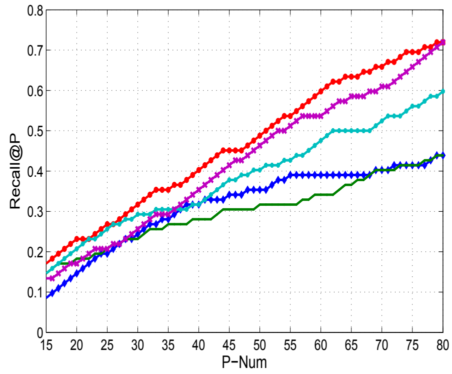
Figure 4.5 shows the technical performance of precision and recall by setting three different window sizes, where the horizontal axis (P-Num) stands for the number of predicted trading weeks in financial crisis, and the vertical axis represents the values of technical measures. We can see that the CHMM based approaches outperform the other approaches at any window sizes. For instance, in Figure 4.5 (c), when P-Num=45, the precision of CHMM-Logistic is 90%, 40% better than Logistic. In addition, recall represents the probability that a crisis is retrieved. And Figure 4.5 (b)(d)(f) show the CHMM based approaches achieve higher recall than observation based models with any P-Nums.

4.5 Summary

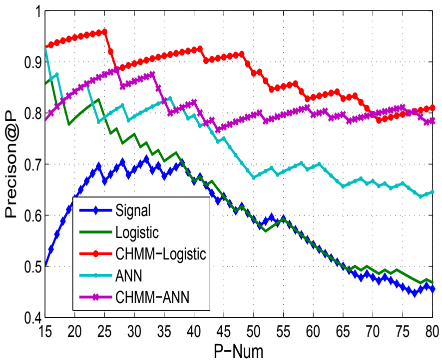
In this chapter, we propose a new financial crisis forecasting framework based on the *CMSA*. In particular, CHMM is employed as a concrete *CSSM* to capture the coupled relations of hidden market states in various markets. We evaluate our approach and other comparative methods over the commodity, stock and interest markets. The results show the clear advantage of our *CMSA* based approach against the state-of-the-art observation based methods with a significant improvement. Obviously, *CMSA* is applicable to many other areas (e.g. currency crisis forecasting, financial market movements prediction) related to cross-market analysis.



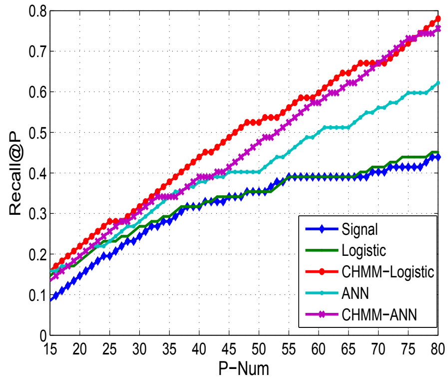
(a) Precision(k=2)



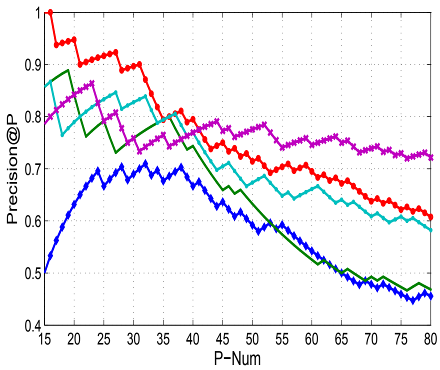
(b) Recall(k=2)



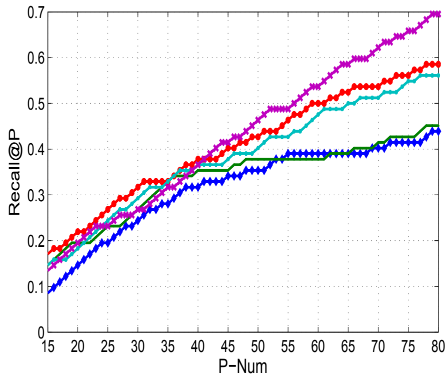
(c) Precision(k=3)



(d) Recall(k=3)



(e) Precision(k=4)



(f) Recall(k=4)

Figure 4.5: Technical Performance of Five Approaches for Financial Crisis Forecasting (k denotes window size) 77

Chapter 5

Market Trends Forecasting via Coupled Cross-Market Behavior Analysis

Since the availability of large profits can result in people leading a comfortable life, the financial markets represent the most attractive investment choices. However, forecasting market trends is a challenging task due to volatile, complex and non-linear characteristics. In this chapter, we investigate a novel approach to predict one market's movements through analyzing its hidden coupling relationships with different types of other related financial markets. This motivates the research on cross-market behavior analysis, involving coupled behaviors (intra-coupling and inter-coupling) and hidden relationships across multiple markets, in addition to the non-linear and dynamic features of market dynamics.

A framework of coupled cross-market behavior analysis (CCBA) is proposed in this chapter to explore the nonlinear and dynamic hidden couplings between indicators in different financial markets, to forecast the trends of one market. Through capturing the coupled market behaviors, a CHMM is built to infer a market trend by forecasting its probability distributions. The experimental results from nine years of real financial market data show that the proposed approach produces more investment profits as compared to other baselines.

5.1 Background and Overview

The subprime mortgage crisis which occurred in the US in 2007 has caused a chain of disruptive effects on global financial markets. A higher accuracy of financial market movements forecasting is of paramount concern. This is ever important for traders who have been struggling in the new financial environment.

However, the effective forecasting of market trends is notoriously difficult and presents some major challenges. Firstly, financial market movements are highly complex behaviors, with many interactive factors, especially the complex interactions across different markets, which jointly contribute to market dynamics. In the literature, the existing work tends to infer a market's trends solely on its own historical data, which do not capture the impact of other related market factors. This shows not operable (de Souza e Silva, Legey & de Souza e Silva 2010), and triggers the consideration of combining indicators from multiple markets to forecast market trends. Secondly, the indicators from different financial markets are often highly coupled with each other, forming *cross-market* interactions and effect. For example, a stock market's trends are affected by other related commodity and interest markets. This may build a strong transfer effect from one market to another through complex coupling relationships, as evidenced by the 2008 global financial crisis (Longstaff 2010). Such complex interactions are driven by hidden behaviors that are not observable directly from market indexes (Chan et al. 2011). In addition, financial markets are dynamic and nonlinear. For example, the $S\&P$ 500 stock market index dropped from 1267.38 in July to 896.24 in November during the 2008 global financial crisis, which was approximately 10% down per month. This makes it difficult for approaches with the linear assumption (e.g. logistic model) to capture financial market movements. These aspects form the need and challenges of *cross-market behavior analysis*.

However, very limited work has been reported on such cross-market behavior analysis, for such problems as market trends forecasting. This urges the need to develop approaches for effectively catering for the dynamic non-linear characteristics, and more importantly capturing the complex hidden couplings among multiple related financial markets. To this end, we propose a new perspective: *coupled cross-market behavior analysis* (CCBA), and the corresponding tools, to forecast financial market movements. Here coupled cross-market behaviors refer to activities of different related markets, such as stock, commodity and interest markets that are coupled with each other.

The forecasting process consists of the following steps. We firstly introduce the CCBA to analyze the complex hidden coupling relationships between various markets. Secondly, two mapping processes are introduced to formalize the forecasting issue specifically. The first mapping relation is formed from CCBA to the Coupled Hidden Markov Model (CHMM), which is used to model multiple series of couplings across markets. The second mapping process is from CHMM to specific financial market movements forecasting. Finally, we will infer a market trends by forecasting its probability distribution through capturing the hidden coupled cross-market behaviors via a CHMM.

The remainder of this chapter is organized as follows. In Section 5.2, the coupled market behavior problem is illustrated by a case study, CCBA is introduced, and then the corresponding problem is defined. Section 5.3 describes our proposed framework, including the two mapping processes mentioned above. The effectiveness of CCBA is shown in Section 5.4 with intensive experiments. We conclude this work and address future work in Section 5.5.

5.2 Problem Statement

5.2.1 A Case Study

The 2008 global financial crisis shows the linkages existing in different financial markets (Chevallier 2012). Here the coupling relations are verified through a quantitative perspective, by using typical market indexes of three major financial markets (Commodity market: The Gold price and The WTI Crude Oil Futures Price; Stock market: The S&P 500 index and Dow Jones Industrial Average (DJIA); Interest market: The TED Spread and The BAA Spread).

The coupling relationships among the three types of markets are illustrated clearly in Table 5.1 from quantitative perspective. Table 5.1 uses the Pearson correlations to describe the relations between the indicators in the three markets and the data is from January 1990 to December 2013 (including financial crisis and non-crisis periods). We can see clearly that the coefficients are very significant through the whole period, namely there are strong correlations between the three markets.

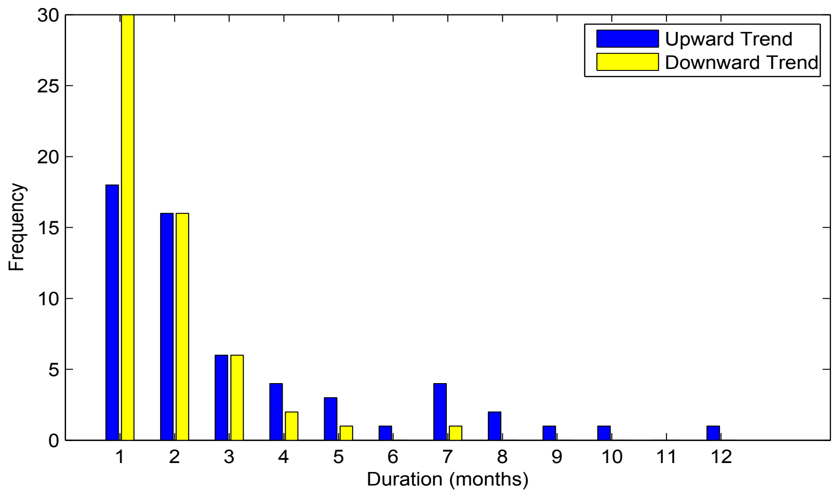


Figure 5.1: Frequency of the DJIA Trends Duration

Table 5.1: Correlations between Indicators in Three Types of Markets (1990-2013)

		BAA	WTI OIL	S&P 500	GOLD	DJIA	TED
BAA	Pearson Correlation	1	.406**	.292**	.465**	.356**	.190**
	Sig.(2-tailed)		.000	.000	.000	.000	.001
WTI OIL	Pearson Correlation	.406**	1	.627**	0.867**	.717**	.044
	Sig.(2-tailed)	.000		.024	.000	.000	.464
S&P 500	Pearson Correlation	.282**	.827**	1	.443**	.982**	.019
	Sig.(2-tailed)	.000	.000		.000	.000	.756
GOLD	Pearson Correlation	.455**	0.857**	0.443**	1	.550**	-.117*
	Sig.(2-tailed)	.000	.000	.000		.000	.048
DJIA	Pearson Correlation	.356**	.717**	.982**	.550**	1	-.019
	Sig.(2-tailed)	.000	.000	.000	.000		.746
TED	Pearson Correlation	.190**	.044	.019	-.117*	-.019	1
	Sig.(2-tailed)	.001	.464	.756	.046	.746	

** Correlation is significant at the 0.01 level(2-tailed), * Correlation is significant at the 0.05 level(2-tailed).

In addition, as illustrated by the the DJIA index, Figure 5.1 demonstrates the time duration of the upward and downward trends. It can be observed that the upward trends tend to last longer than the downward trends, which means that the behaviors of the two trends are different in stock market.

All the above discussions support our assumption that there are complex hidden coupling relationships between various financial markets. Also, the difference between the upward and downward trends in financial market brings about the possibility to distinguish them.

5.2.2 Coupled Cross-Market Behavior Analysis

Coupled cross-market behaviors (CCB) refer to the activities from multiple markets with inter- and intra- relations between markets. Suppose there are I markets $\{M_1, M_2, \ldots, M_I\}$, a market M_i undertakes J market behaviors $\{MB_{i1}, MB_{i2}, \ldots, MB_{iJ}\}$. Then a Market Behavior Feature Matrix $FM(MB)$ is defined as follows:

$$FM(MB) = \begin{pmatrix} MB_{11} & MB_{12} & \ldots & MB_{1J} \\ MB_{21} & MB_{22} & \ldots & MB_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ MB_{I1} & MB_{I2} & \ldots & MB_{IJ} \end{pmatrix}$$

Then the intra-couplings within each market is the relationships within one row of the above matrix, while how the behavior interactions between different markets are embodied through the different rows of $FM(MB)$, indicated as inter-couplings.

Definition 5.1 *Coupled Cross-Market Behaviors (CCB) refer to behaviors* $MB_{i_1j_1}$ *and* $MB_{i_2j_2}$ *that are coupled in terms of relationships* $f(\theta(\cdot), \eta(\cdot))$*:*

$$CCB = MB_{i_1j_1} \odot MB_{i_2j_2} | f(\theta(\cdot), \eta(\cdot)) \tag{5.1}$$

where $\odot$ *denotes the coupled interactions between different market behaviors* MB*.* $\theta(\cdot)$ *is the coupling function indicates the market* M_i*'s behaviors* MB_{ij}

are intra-coupled, while $\eta(\cdot)$ *represents the inter-coupling relations between different market's behaviors.*

Definition 5.2 *Coupled Cross-Market Behavior Sequences (CCBS). Suppose* MB *is partitioned into* m *coupled market behavior sequences, then*

$$\Phi(CCBS) = \Phi(MBS_i) * \Phi(MBS_j)|f(\theta(\cdot), \eta(\cdot)) \tag{5.2}$$

where $*$ *denotes the coupled interactions between different market behavior sequences* MBS*, which can be represented by* r_{ij} *which is a set for all* m *market behavior sequences* $((1 \leq i, j \leq m))$*. So if* $r_{ij} = \varnothing$*, there is no coupling relationship in market behavior sequences* $\Phi(MBS_i)$ *and* $\Phi(MBS_j)$*, the bigger the* r_{ij}*, the stronger relationship between the two market behavior sequences* $\Phi(MBS_i)$ *and* $\Phi(MBS_j)$*.*

Definition 5.3 *Coupled Cross-Market Behavior Analysis (CCBA) is to build the objective function* $g(MB)$ *under the condition that market behaviors are coupled with each other by coupling function* $f(\cdot)$*, and satisfy the following conditions:*

$$f(\cdot) ::= f(\theta(\cdot), \eta(\cdot)) \tag{5.3}$$

$$g(MB_{i_1 j_1})|f(\cdot) \geq g(MB_{i_2 j_2})|f(\cdot) \tag{5.4}$$

The above discussion defines the basic concept of CCBA, and the vector-based market behavior representation show us the road map from understanding to modeling CCBA in the real financial market. In the following section, we discuss the specific mapping and modeling issues.

5.2.3 Problem Formalization

In the above, the forecasting problem can be formalized as follows: the coupling function $f(\cdot)$ here is used to capture the complex interactions between different financial markets, and the objective function $g(MB_c)$ is build to forecast the possibilities of two trends (upward and downward) for a market c at the trading day $t+1$, where $MB_c = \{MB_{cu}, MB_{cd}\}$, namely $g(MB_{cu})$

represents the possibility of upward trends within market c while $g(MB_{cd})$ evaluates the downward trends. So if at time t,

$$g_{t+1}(MB_{cu}) \mid f_t(\cdot) \geq g_{t+1}(MB_{cd}) \mid f_t(\cdot) \tag{5.5}$$

then time $t+1$ is a upward trend, and otherwise a downward trend.

The key task of market trends forecasting then is to build a proper model to determine the specific coupling function $f(\cdot)$ and the corresponding objective function $g(MB_c)$. Below, CHMM is explored to capture the complex coupling relationships and nonlinear dynamics of multiple markets, with an objective function to forecast the following trends based on CHMM.

5.3 The Proposed Approach

This section describes the methodology we developed to forecast financial market trends. It is composed of four different steps, all of which are, individually, illustrated hereafter.

5.3.1 Indicator Selection

As mentioned in Section 5.2, each financial market owns more than one indicator. In CHMM, we use one Markov chain to represent a market, and here we select one indicator for each market which has higher correlations with other markets. This is because our focus is on the coupling relationships among various markets, hence higher correlation with other markets encloses a stronger discriminative power.

Suppose there are I markets, and each market owns N indicators. $MIC_{i_1j_1}$ refers to the correlations of market indicator $MI_{i_1j_1}$ with indicators in other markets, where $(i_1 \neq i_2)$.

$$MIC_{i_1j_1} = \sum_{i_2=1}^{I} \sum_{j_2=1}^{N} corr(MI_{i_1j_1}, MI_{i_2j_2}) \tag{5.6}$$

where $corr(\cdot)$ is the Pearson correlation coefficient of the two market indicators.

5.3.2 CHMM-based Coupled Market Behavior Modeling

The modeling can be divided into two mapping process: one from CCBA to CHMM, and another from CHMM, to a specific market trends forecasting problem. The first mapping process is to formalize the general analysis to a specific tool, and the latter is to resolve the forecasting problem with the tool.

A. Mapping Process from CCBA to CHMM

In this chapter, we illustrate our approach via three types of markets: stock, commodity and interest markets. Correspondingly, there are three Markov chains, namely HMM-S enclosing the stock market sequence $\Phi(MBS_S)$, HMM-C capturing the commodity market sequence $\Phi(MBS_C)$, and HMM-I for the interest market sequence $\Phi(MBS_I)$. The representation of coupled market behavior sequences in Equation (5.2) shows the possibility of using CHMM to model CCBA. The mapping from CCBA to CHMM works as following:

$$CCBA \rightarrow CHMM\, modeling$$

$$f(\theta(\cdot), \eta(\cdot)) \rightarrow \Omega(A, B, R, \pi) \tag{5.7}$$

$$\Phi(\cdot)|transition \rightarrow A(Transitional\, probability) \tag{5.8}$$

$$\Phi(\cdot)|observation \rightarrow B(Observation\, probability) \tag{5.9}$$

$$r_{ij} \rightarrow R(Coupling\, coefficient) \tag{5.10}$$

$$\Phi(\cdot)|prior \rightarrow \pi(Prior\, Probability) \tag{5.11}$$

$$g_{t+1}(MB_{cu}) \mid f_t(\cdot) \rightarrow P(MB_{cu} \mid \Omega) \tag{5.12}$$

$$g_{t+1}(MB_{cd}) \mid f_t(\cdot) \rightarrow P(MB_{cd} \mid \Omega) \tag{5.13}$$

where $i, j \in \{stock\ market, commodity\ market, interest\ market\}$, r_{ij} represents the coupling coefficient of two markets i and j.

B. Mapping Process from CHMM to Market Trends Forecasting

In the above, we have mapped CCBA to a CHMM model, so then the issue is how to decode the forecasting issue into CHMM, namely finding the mapping process from CHMM to stock market forecasting. Suppose there are H hidden state in an HMM, which are denoted as $Z = \{Z_1, Z_2, \ldots, Z_j, \ldots, Z_H\}$, where Z_j is an individual state, and the state at time t is denoted as z_t. There are T observations in an HMM, provided as $O = \{O_1, O_2, \ldots, O_i, \ldots, O_T\}$, O_i is an individual observation symbol, while o_t is the observation symbol at time t. The mapping from CHMM to trends forecasting in market $c \in \{S, C, I\}$ works as follows:

$$CHMM \rightarrow Market\ Forecasting\ Mapping$$

$$A(Transitional\, probability) \rightarrow P(z_{t+1}^c = Z_j \mid z_t^{c'} = Z_j') \tag{5.14}$$

$$B(Observation\, probability) \rightarrow P(O_t^c \mid z_t^c = Z_j) \tag{5.15}$$

$$R(Coupling\, coefficient) \rightarrow r(c, c') \tag{5.16}$$

$$\pi(Prior\ Probability) \rightarrow \pi^c \tag{5.17}$$

$$P(MB_{cu} \mid \Omega) \rightarrow P(z_{t+1}^c = up \mid O_{1:t}^{S,C,I}, \Omega) \tag{5.18}$$

$$P(MB_{cd} \mid \Omega) \rightarrow P(z_{t+1}^c = down \mid O_{1:t}^{S,C,I}, \Omega) \tag{5.19}$$

where $(c, c' \in \{S, C, I\}) \wedge (c \neq c')$, $H = 2$, $1 \leq j, j' \leq H$, namely there are two hidden states: $\{upward,\ downward\}$. $P(z_{t+1}^c \mid O_{1:t}^{S,C,I}, \Omega)$ is to forecast movement in market c at time $t+1$ based on the observations of the indicators from the three markets in the time interval $[1, t]$ $(O_{1:t}^{S,C,I})$. Then if the upward trend probability at time $t+1$ $(P(z_{t+1}^c = up \mid O_{1:t}^{S,C,I}))$ is larger than the downward trend probability at time $t+1$ $(P(z_{t+1}^c = down \mid O_{1:t}^{S,C,I}))$, time $t+1$ is an upward trend in market c, otherwise the downward trend.

In above, we incorporate the CCBA into market trends forecasting based on CHMM. Below, we discuss the specific process for the forecasting, namely how to calculate the $P(z_{t+1}^c = up \mid O_{1:t}^{S,C,I}, \Omega)$ and $P(z_{t+1}^c = down \mid O_{1:t}^{S,C,I}, \Omega)$.

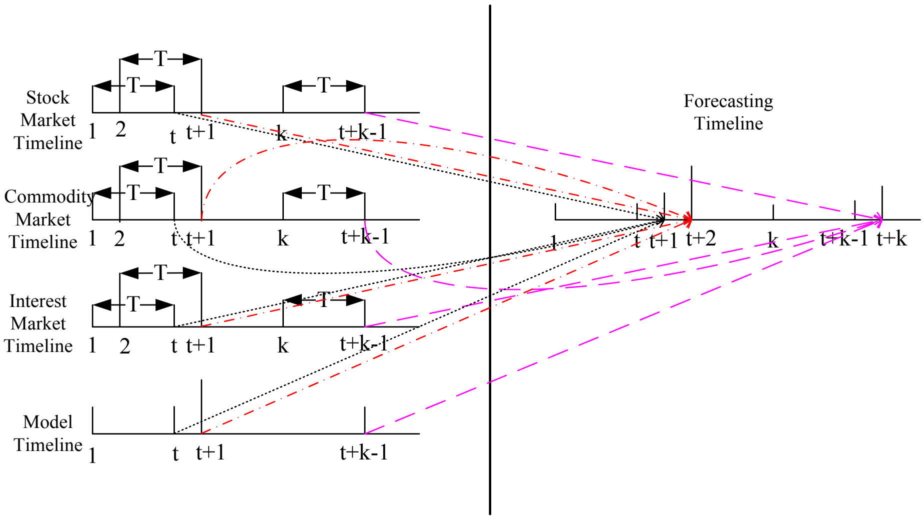


Figure 5.2: CHMM-based Forecasting Process

5.3.3 Market Forecasting Process

Figure 5.2 illustrates the general framework of the proposed forecasting process. At each time interval $[1, t]$ (the length here is T) , the first step is to train CHMM using the most recent T observations in three markets $O_{1:T}^{S,C,I}$, and then we can obtain the current model $\Omega^t(A, B, R, \pi)$ (Every T days, the parameters in CHMM will be re-estimated using the most recent T observations). Then, based on the trained model Ω and the T observations $O_{1:T}^{S,C,I}$, the CHMM gives the probability distribution for the following movements of market c at time $t+1$. Then we can obtain the market trend at time $t+1$ through comparing the probabilities of forecasted upward trend $(P(z_{t+1}^c = up \mid O_{1:t}^{S,C,I}))$ and the downward trend $(P(z_{t+1}^c = down \mid O_{1:t}^{S,C,I}))$ at time $t+1$. k is the number of forecasting points, namely the number of sliding windows. The specific deviation steps are as follows[1]:

[1]Here we omit the Ω^t, rather fixing it for simplicity. The conditional independence properties are used (Jordan 2002).

$$
\begin{aligned}
P(z_{t+1}^c = up \mid O_{1:t}^{S,C,I}) &= \sum_{z_t} p(z_{t+1}^c = up, z_t^{S,C,I} \mid O_{1:t}^{S,C,I}) \\
&= \sum_{z_t} p(z_{t+1}^c = up \mid z_t^{S,C,I}, O_{1:t}^{S,C,I}) p(z_t^{S,C,I} \mid O_{1:t}^{S,C,I}) \\
&= \sum_{z_t} p(z_{t+1}^c = up \mid z_t^{S,C,I}) p(z_t^{S,C,I} \mid O_{1:t}^{S,C,I}) \\
&= \sum_{z_t} p(z_{t+1}^c = up \mid z_t^{S,C,I}) \frac{p(z_t^{S,C,I}, O_{1:t}^{S,C,I})}{p(O_{1:t}^{S,C,I})}
\end{aligned} \tag{5.20}
$$

Equation (5.20) can be evaluated with the following steps: firstly, we run a forward recursion, from which we obtain $P(z_t^{S,C,I}, O_{1:t}^{S,C,I}) = \prod_{c'} \alpha(z_t^{S,C,I})$ and $P(O_{1:t}^{S,C,I}) = \prod_{c'} \sum_{z_t} \alpha(z_t^{S,C,I})$, where $c' \in \{S, C, I\}$. After this, we compute the final summation over z_t. Then we obtain the probabilities of upward and downward trends at time $t+1$. Once the observation at time $t+1$ is observed, we run the α recursion forward to the next step in order to predict the subsequent movement at time $t+2$, and this can carry forward indefinitely. We derive the corresponding parameters $\omega \in \Omega(A, B, R, \pi)$ by the forward-backward procedure similar to (Zhong & Ghosh 2001). The calculation of $P(z_{t+1}^c = down \mid O_{1:t}^{S,C,I})$ follows the similar steps.

One thing that needs to mention here is $p(z_{t+1}^c \mid z_t^{S,C,I})$, it is the state transition probability illustrated in Equation (5.14), and it is computed here by

$$
p(z_{t+1}^c = up \mid z_t^{S,C,I}) = \sum_{c'} r_{c'c} \cdot p(z_{t+1}^c = up \mid z_t^{c'}), c' \in \{S, C, I\} \tag{5.21}
$$

here $r_{c'c}$ is the coupling coefficient which measures the coupling relationships from market c' to market c, namely how $z_t^{c'}$ affects z_{t+1}^c.

5.3.4 The Forecasting Algorithm

With the probability distributions for the next trend of a market, here we further determine the specific follow-up trend of the market c through comparing the probability of the following two trends (upward and downward trends). The corresponding algorithm is described in Algorithm 5.1. The input is the observations of three financial markets $O^i(O^i = O^{S,C,I}_{i:t+i-1})$, and the time interval is T. It is a loop process to obtain w times of forecasting, namely the number of forecasting windows is w. We firstly train the model $\Omega^i(1 \leq i \leq w)$ according to the observations $O^i(O^i = O^{S,C,I}_{i:t+i-1})$, then we forecast the probability of two trends: upward or downward, respectively. The output of the algorithm includes two sets: upward set UWS and downward set DWS, and the total objects in the two sets is w.

Algorithm 5.1 Financial Market Trends Forecasting via CCBA

Require: An observation set $\{O^1, O^2, \ldots, O^w\}$

for O^i in the observation set **do**
  Train the model Ω^i on the specific O^i;
  Compute the probability of the two trends given the model Ω^i and observations O^i at time $t+i$, respectively:
  $P_{t+i}(z^c_{t+1} = up \mid O^i, \Omega^i)$ and $P_{t+i}(z^c_{t+1} = down \mid O^i, \Omega^i)$
  if $P_{t+i}(z^c_{t+1} = up \mid O^i, \Omega^i) \geq P_{t+i}(z^c_{t+1} = down \mid O^i, \Omega^i)$ **then**
    $trend\, at\, time\, t+i \rightarrow UWS$
  else
    $trend\, at\, time\, t+i \rightarrow DWS$
  end if
end for
return An upward trends set UWS; A downward trends set DWS

5.4 Experiments

In this section, we evaluate the proposed method on real financial markets data, and compare the results with other related methods.

5.4.1 Experimental Settings

A. Data Preparation

This chapter aims to forecast financial market movements by considering coupling relationships between the related financial markets. Thus, the data set of interest is the indexes of history prices in different financial markets. As discussed in Section 5.3, we only choose one indicator for each market, according to its correlations with other markets. Based on Equation (5.6), the selected indicators are: DJIA for stock market, WTI Crude Oil price for commodity market and the BAA Spread for the interest market.

The data set includes weekly closing prices from January 2005 to December 2013, obtained from the Economic Research (http://research.stlouisfed.org/), and the prices of each market are decoded into [0, 1] based on $PI'^c = (PI^c - PI^c_{min})/(PI^c_{max} - PI^c_{min})$, here PI^c_{max} and PI^c_{min} are the maximum price and minimum price in market c, respectively. Indicators in different markets may appear on different trading days, we delete those days on which some market data is missing and only choose the days with trading data from all financial markets. The total number of dataset is 470 observations.

B. Parameter Specification

The CHMM parameter specification involves two steps. The first defines the CHMM elements, and the second deals with the CHMM initial values.

- Specification of the CHMM Elements. As mentioned in Section 5.3, three HMM sequences represent the three financial markets, so the number of Markov chains is three. In this chapter, the number of states is two (representing the physical significance of upward and downward). In addition, the size of the forecasting window, namely the forecasting interval T, was chosen which is neither too long (because of the uncertainty), nor too small (because of the high volatility of prices). So, according to the domain knowledge and several tests in the experiments, here we set T equal to 30.

- CHMM Initial Parameter Setting. The CHMM parameter estimation is done using the EM algorithm. We also follow the common idea of using random starting values for the parameters and starting the algorithm from several different starting points and then selecting a better one (Harshman & Lundy 1994). Here the initial parameter values of π and A are set by following the random selection method.

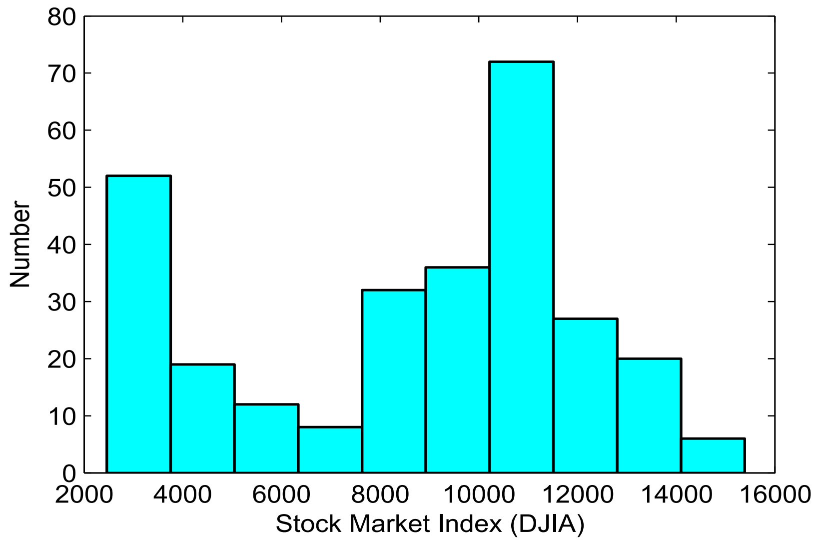


Figure 5.3: The Distribution of Stock Market (DJIA)

The parameter which needs an initial value here is B (Observation probability). To this end, we need to find the corresponding number of the mixture component. This is because, in an infinite Gaussian mixture model, how to find the right number of the mixture components is an important and difficult issue, and the right number will help obtain promising results (Rasmussen 1999). As shown in Figure 5.3 (here we use the stock market as an example), the distribution of stock market is a mixture Gaussian, and we need to find the number of the mixture component. Based on the former research, as one of the most popular Bayesian nonparametric models (Teh 2010), the Dirichlet

Process (DP) is used in infinite mixture models. Here, we use DP to find the correct numbers of mixture components in three Markov chains for the three financial markets. The results reveal that the numbers of mixture components in the three Markov chains are all equal to two.

C. Comparative Methods

To evaluate our approach, we take the following typical methods which are either typically used in financial markets or directly address market forecasting

- ARIMA. This is a statistical method for analyzing and building a forecasting model which best represents a time series by modeling the correlations in the data. Taking advantage of its strictly statistical approach, the ARIMA approach only requires the prior data of a time series to generalize the forecast. The form of an $ARIMA(p, d, q)$ model is as $(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q} \theta_i L^i)\varepsilon_t$,where p is the number of autoregressive terms, d is the number of non-seasonal differences, q is the number of lagged forecast errors, ϕ_i are the regression coefficients, θ_i are the parameters of the moving average part, ε_t are error terms and L is the lag operator. New forecasts can be made for the process $Y_t = (1 - L)^d X_t$, using a generalization of the method of autoregressive forecasting.

- Logistic. The input variables are indicators from the three different global financial markets, and the parameter estimation is following M-LE. In addition, here we treat market upward and downward as two categories.

- ANN (Hyup Roh 2007). Following the back-propagation algorithm, we train this model with indicators from the three markets.

- Buy & Hold Strategy. According to the trading strategy, an investor buys an asset and holds it for a long period of time. In this chapter, the investor uses the strategy to buy the market index in the beginning and maintains this position during the whole trial.

- HMM (Zhong & Ghosh 2001). This is a sub-model of CHMM which simply models one type of market behavior without considering the complex couplings between different financial markets.

5.4.2 Evaluation Metrics

A. Technical Perspective

we use following technical metrics:
Accuracy$=\frac{TN+TP}{TP+FP+FN+TN}$.
Precision$=\frac{TP}{TP+FP}$.
Recall$=\frac{TP}{TP+FN}$.
where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively. We treat the upward trend as the positive class here.

B. Business Perspective

Rate of Return (ROR) is the ratio of money gained or lost on an investment relative to the amount of money invested.
Rate of Return (ROR)$=\frac{Final\ Captical-Initial\ Capital}{Initial\ Capital}$.

5.4.3 Experimental Results

A. Technical Performance

In this part, we present results in stock and commodity markets. The testing data is from January 2006 to December 2013, includes a crisis period (2007-2009) and a non-crisis period (2010-2013). This arrangement aims to disclose the model performance against different market situations with distinct interactions.

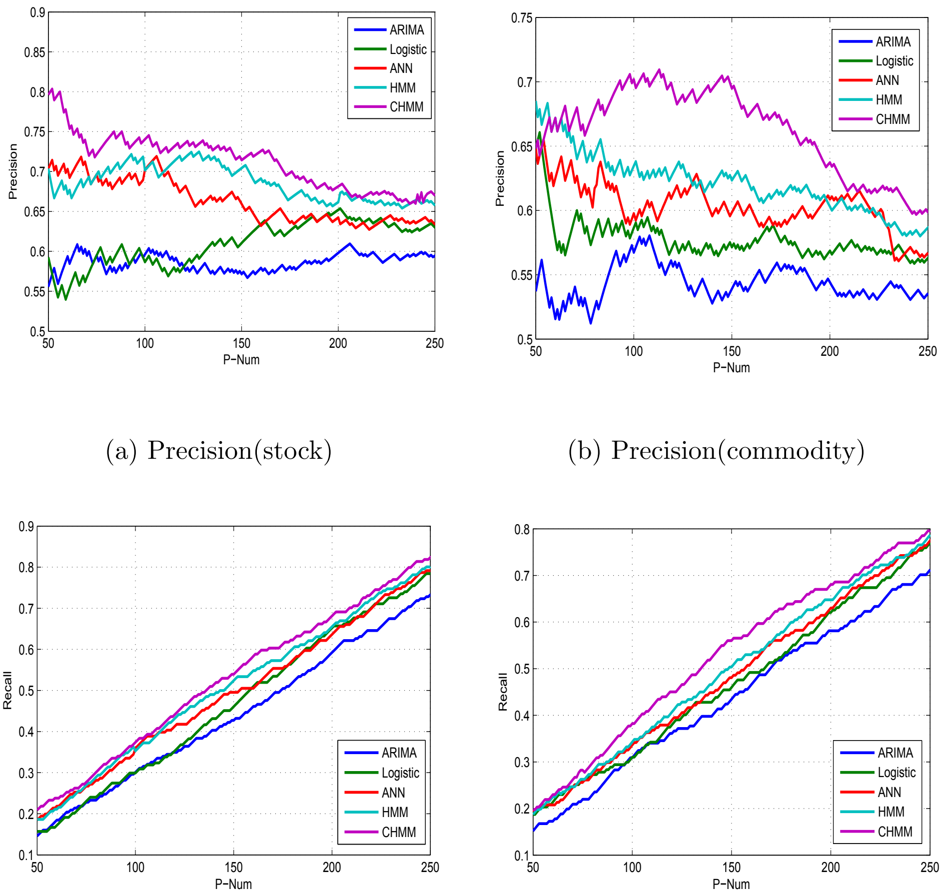
Table 5.2: Accuracy of Comparative Methods in Stock and Commodity Markets

Year	Stock Market					Commodity Market				
	ARIMA	Logistic	ANN	HMM	CHMM	ARIMA	Logistic	ANN	HMM	CHMM
01/06-12/06	0.5962	0.5962	0.6154	**0.6346**	**0.6346**	0.5962	0.6154	0.6154	**0.6346**	0.6154
01/07-12/07	0.6154	0.6346	0.6346	0.6346	**0.6538**	0.5769	0.5962	0.5962	0.5962	**0.6154**
01/08-12/08	0.5769	0.5962	0.5962	0.6154	**0.6538**	0.5000	0.5769	0.5962	0.6154	**0.6346**
01/09-12/09	0.6154	0.5769	0.6154	0.6154	**0.6346**	0.5962	0.6154	**0.6346**	0.6154	**0.6346**
01/10-12/10	0.6346	0.6154	0.6346	0.6346	**0.6538**	0.5385	0.5962	0.5962	0.6154	**0.6731**
01/11-12/11	0.5000	0.6154	0.6346	0.6538	**0.6731**	0.5962	0.5577	0.6154	0.6346	**0.6538**
01/12-12/12	0.5962	0.5962	0.6154	0.6154	**0.6538**	0.5769	0.5962	0.6154	0.6154	**0.6346**
01/13-12/13	0.5962	0.6154	0.6154	0.6346	**0.6538**	0.5740	0.5962	0.6154	0.5962	**0.6731**
01/06-12/13	0.5926	0.6057	0.6203	0.6303	**0.6514**	0.5698	0.5937	0.6171	0.6156	**0.6386**

The results of Accuracy is reported in Table 5.2, including yearly and whole period accuracy. From the table we can find that the baseline method ARIMA does not perform well, this is because ARIMA is often failing with dynamic data since it has the stationary assumption. More importantly, it pays no attention to the underlying complex hidden interactions between the different markets. For the similar reason, the Logistic, ANN and HMM models do not perform very well. Note that ANN outperforms the Logistic approach, the main reason here is that Logistic approach is under a linear assumption, but the financial market, especially the hidden couplings, are not linear. Interestingly, the outcomes of ANN and HMM conflict with each other, as shown in the whole testing period, HMM performs better in stock market while ANN does a better forecasting in commodity market. This may be because, ANN considers the different market indicators in forecasting while HMM overlooks it, but HMM can capture the hidden market behaviors which is a key factor in market movements while ANN cannot handle it. However, both of them overlook the complex hidden couplings between various markets.

Our CHMM approach outperforms all of the baselines regardless of yearly data or in the whole period. This can be interpreted as follows: firstly, CHMM learns intra- and inter- couplings between different markets, which serve as the key factors driving market dynamics; secondly, unlike those methods that predict market movements directly from the observations, CHMM builds an architecture to learn the coupled hidden market behaviors which removes the vulnerabilities of observations.

Figure 5.4 plots the precision and recall of all comparative approaches in stock and commodity markets, where the horizontal axis stands for the number of predicted trading weeks in upward trends, and the vertical axis represents the values of technical measures. We can see that our CHMM outperforms all other comparative methods. For example, the precision improvement in Figure 5.4 (b) is as high as 15% against the ARIMA approach, and around 10% against the ANN method when P-Num equals to 150. Figure 5.4 (c) (d) show the CHMM achieves higher recall than other models with any number of predicted trading weeks.



(a) Precision(stock)

(b) Precision(commodity)

(c) Recall(stock)

(d) Recall(commodity)

Figure 5.4: Precision and Recall of Comparative Methods in Stock and Commodity Markets

B. Business Performance

We also compare the business performance of all comparative approaches in stock market. The results are reported in Table 5.3 and Figure 5.5.

Tables 5.3 illustrates ROR yearly, as the ROR indicator is important for

Table 5.3: ROR of Comparative Methods in Stock Market

Year	ROR					
	ARIMA	Logistic	ANN	HMM	CHMM	Buy & Hold
01/06-12/06	0.1590	0.1178	0.0150	0.2504	**0.2714**	0.2441
01/07-12/07	-0.0283	-0.0288	-0.0401	0.0390	**0.0589**	-0.0736
01/08-12/08	-0.0805	-0.0137	0.2141	0.1778	**0.3447**	-0.3365
01/09-12/09	**0.1976**	0.1742	0.1955	0.1850	0.1593	0.3017
01/10-12/10	0.0885	0.0412	0.0585	**0.0847**	0.0520	0.0412
01/11-12/11	0.1330	0.1277	0.1418	0.0450	**0.1592**	0.1339
01/12-12/12	-0.0678	**0.1328**	-0.0733	0.0088	0.0940	0.0137
01/13-12/13	-0.0410	0.1135	0.0676	0.1380	**0.2011**	0.1135

investors to validate the actionability of outcomes in real financial markets. The table shows that our approach performs much better than other approaches. For instance, our approach is with the highest ROR with a gain of about 24% compared to the ARIMA method, and 6% over the best performing HMM model in 2013. In particular, through the financial crisis period (2007-2009), our approach can help investors to make more stable profits compared to other approaches. This means that our approach can better capture those bigger losing trends, which can help investors to reduce loss.

Figure 5.5 shows an investor's wealth evolution by investment following the six approaches from January 2006 to December 2013. We have the following settings for the investment: (1) the initial capital investment is USD100; (2) no new capital will be added thereafter; (3) an investor buys and sells the index according to the trends predicted by each approach (buy when there is an upward forecasting and sell while downward); and (5) the Buy & Hold strategy is used as the benchmark. The following conclusions can be drawn from the figure: the CHMM approach performs best, that is, an investor taking the recommendations from the approach can make profit at $177.8, which represents a return of 170% in eight years (including the financial crisis period). In addition, CHMM-based CCBA has been demonstrated to be a useful tool to help investors to make wiser trading decisions.

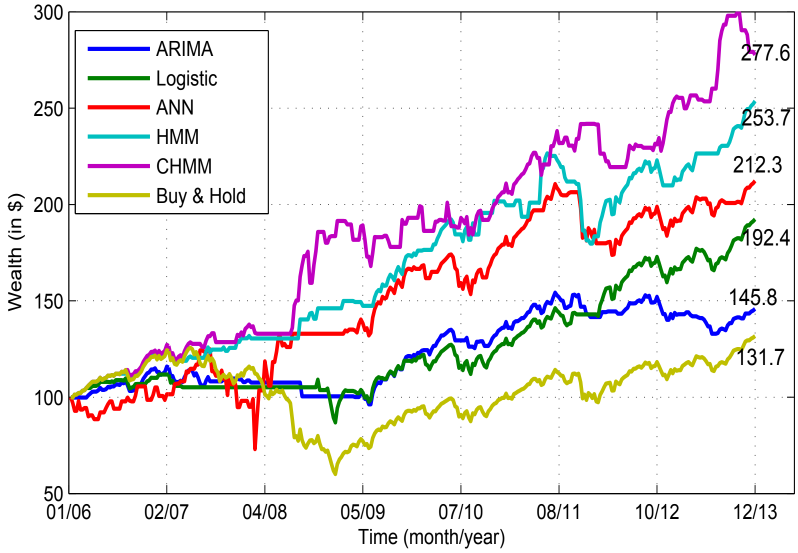


Figure 5.5: Investor's Wealth Evolution in Stock Market

5.5 Summary

In this chapter, we develop a new forecasting approach to capture financial market trends by analyzing the complex coupling relationships between various indicators in different financial markets. A CHMM-based forecasting model is built to analyze the hidden couplings between stock, commodity and interest markets. The results show that the prediction from the proposed method can gain better investment outcomes compared to the state-of-the-art methods, from technique and business perspectives. Future directions include: (1) extending the model to predict real market index prices; and (2) incorporating global financial information(e.g. political news) to detect casual relations among different financial markets.

Chapter 6

Stock Market Trend Forecasting via Hierarchical Cross-market Behavior Analysis

In this chapter, we forecast stock market trends through revealing the hierarchical cross-market behaviors hidden in the hierarchical market structure of different markets in various countries. Here, the hierarchical cross-market behaviors refer to the two layers of couplings across multiple financial markets. The Layer-1 coupling represents the interactions between markets in the same country (intra-country market level), the Layer-2 coupling refers to the interactions between different markets in one country (inter-country market level).

A case study is provided to illustrate the existence of coupled relations with two financial markets in 13 countries. After this, a framework of hierarchical cross-market behavior analysis (HCBA) is proposed to predict stock market trends, by exploring the above two layers of couplings. Formally, a Multi-layered Coupled Hidden Markov Model (MCHMM) is built to capture the complex hierarchical cross-market behaviors. Following this, the movements of a stock market in a target country can be inferred by forecasting its price return probabilities with the proposed MCHMM. Experimental results on 10 years of data from the stock market and currency market in 13 countries show that our proposed approach outperforms other state-of-the-art approaches.

6.1 Background and Overview

An increasing number of researchers and practitioners recognize the need and challenge of exploring coupling relationships (Ma, Wei & Huang 2013) between market behaviors of different markets (countries) (for short cross-market behaviors) in forecasting one market's movements. However, cross-market behavior analysis and trend prediction is not a trivial thing. It involves a theoretical challenge, that is to learn the complex coupling relationships hidden in heterogeneous financial variables, for example, those between markets within one country and between various countries. Complex coupling relationships are due to economic, social and other interactions and influence, and are subject to uncertain and evolutionary market dynamics.

Cross-market are coupled for various reasons. Taking two types of markets (stock market, currency market) in two respective countries (USA, China) as an example, as shown in Figure 6.1, the future behavior of the US stock market is not only determined by itself and the US currency market (we call intra-country market coupling), but is also influenced by the Chinese stock market[1](we call inter-country market coupling). This was the situation in the 2008 global financial crisis (Longstaff 2010). Hence, we need to effectively capture the comprehensive couplings between the two markets in US and China in order to properly predict the trends of the US stock market.

Cross-market analysis has become increasingly interesting such as underlying-derivative market relation analysis in finance and economics (Ma et al. 2013).

[1]The US stock market is also indirectly affected by the Chinese currency market, and it can be counted included in the influence from Chinese stock market.

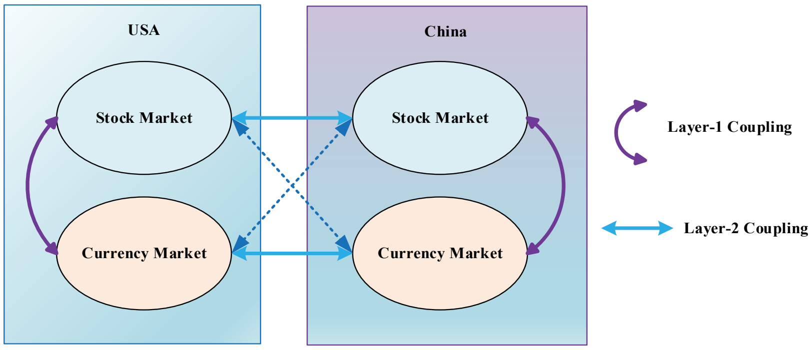


Figure 6.1: An Example of Hierarchical Cross-market

However, very limited research has been conducted in computer science to deeply understand the relationship between financial variables from cross-markets in different countries. The main challenges lie in key aspects including the involvement of heterogeneous financial variables, variables from different markets, complex coupling relationships between behaviors of different markets (Cao et al. 2012), and market dynamics. The couplings may involve intra-market interactions, inter-market interactions, intra-country interactions, and inter-country interactions; if cross-market analysis involves different countries, and we call this *hierarchical cross-market*.

To address these issues, in this chapter, we propose a novel approach: Multi-layer Coupled Hidden Markov Model (MCHMM) for *hierarchical cross-market behavior analysis* (HCBA), to identify systematic market behavioral patterns hidden in cross-market data, namely the complex hierarchical coupling relationships among financial markets in different countries, which are then used to forecast the movements of stock market. Here *hierarchical coupling relationships* refer to two layers of coupled market behaviors: *intra-country market behavior couplings* (Layer-1 coupling) reflecting the interactions between financial markets in the same country, and *inter-country market behavior couplings* (Layer-2 coupling) referring to the influence between the same type of markets across different countries.

The HCBA-based market trend forecasting works as follows. First, HCBA is introduced to formalize the complex coupling relationships between various markets in different countries. Second, two mapping processes are developed to model the forecasting issue. The first mapping relation is formed from HCBA to a MCHMM, in which a CHMM is used to model couplings between markets within a country, and then a MCHMM models the hierarchical market coupling relationships across countries. The second maps a MCHMM to the movement forecasting of a specific stock market. Finally, we infer a stock market trend by estimating its price return probabilities, from capturing the hierarchical coupled market behaviors within a MCHMM. Real financial data in stock and currency markets in 13 major countries is used to evaluate the performance of our MCHMM-based HCBA approach against other classic approaches, including ARIMA, Logistic, ANN and CHMM-based models, from technical and business perspectives. The recommendation results based on the HCBA show the potential for making decisions based on cross-market behavior analysis.

The remainder of this chapter is organized as follows. In Section 6.2, the hierarchical coupled market behavior problem is illustrated by a case study, and the corresponding problem is defined. Modeling framework is introduced in Section 6.3, including HCBA and MCHMM. Forecasting methodology is illustrated in Section 6.4. In Section 6.5, we show the performance of our approach through the corresponding experiments. Conclusions are drawn in Section 6.6.

6.2 Problem Statement

6.2.1 A Case Study

Financial markets are linked with each other, and index changes in one market are influenced by other market dynamics and interactions between markets. The 2008 global financial crisis shows that the cross-market effect transfers not only from one market to another in the same country, but also from one country to others (Chevallier 2012). Here we verify it from a quantitative perspective, by using the market indexes of two major financial markets (stock market and currency market[2]) from 13 major countries[3], as shown in Table 6.1.

Table 6.1: Trading Indexes from 15 Countries

Country	Market	
	Stock Market	Currency Market
USA	^DJI	SDR/USD
Brazil	^BVSP	SDR/BRL
Russia	RTS.RS	SDR/RUB
India	^BSESN	SDR/INR
China	000001.*SS*	SDR/CNY
France	^FCHI	SDR/FRF
U.K.	^FTSE	SDR/GBP
Switzerland	^SSMI	SDR/CHF
Austria	^ATX	SDR/ATS
Germany	^GDAXI	SDR/DEM
Ireland	^ISEQ	SDR/IEP
Netherlands	^AEX	SDR/NLG
Belgium	^BFX	SDR/BEF

Figure 6.2 illustrates the coupling relationships between the US stock market index (^DJI) and other financial market indexes from January 2003 to December 2013. We use the Detrended cross-correlation analysis (DCCA) to quantify the cross-correlations coefficient ρ_{DCCA} between two stationary time series. ρ_{DCCA} is a dimensionless coefficient that ranges between [-1,1]. If two time series are completely cross-correlated (anti cross-correlated) then $\rho_{DCCA} = 1(-1)$, and if there are no cross-correlations between two time

[2]Here we choose Special Drawing Right (SDR) as its numeraire is a potential claim on the freely usable currencies of the International Monetary Fund (IMF) (Jang, Lee & Chang 2011).

[3]Eight developed European countries, USA and BRIC (Brazil, Russia, India and China). BRIC accounted for more than 25% of the world's total GDP according to IMF.

series then $\rho_{DCCA} = 0$. The vertical axis in Figure 6.2 represents the cross-correlation coefficient, while the horizontal axis is the equal length of non-overlapping segments with the time series, more details refer to (Podobnik & Stanley 2007).

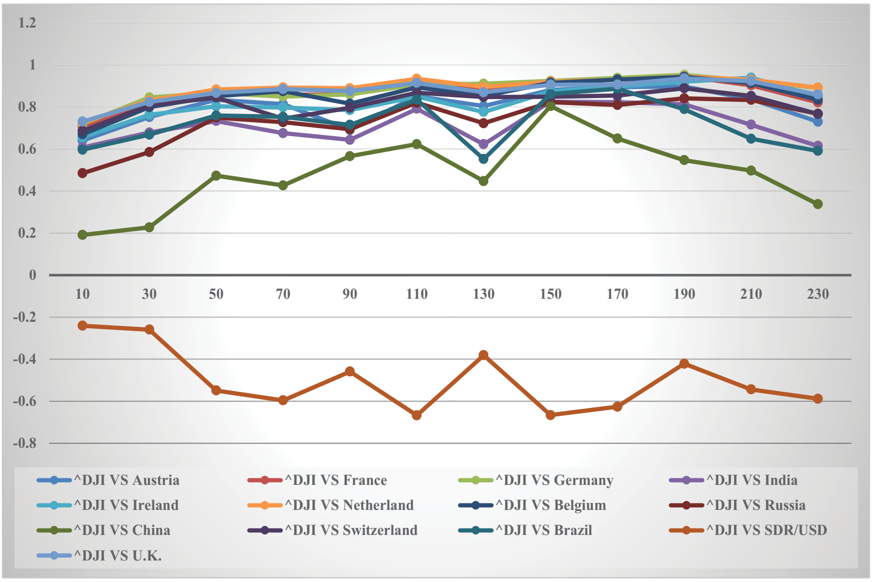


Figure 6.2: Correlations between ^DJI and other Financial Markets

Figure 6.2 shows that there are strong cross-correlations between the US stock market and other financial markets, including the US currency market and stock markets in other 12 countries.

All the above discussions support our assumption that there are complex coupled relationships between not only financial markets in the same country (intra-country market coupling), but also the same markets in different countries (inter-country market coupling), we call this hierarchical cross-market behaviors. Also, the above analysis provides the potential of predicting the movement of a stock market based on exploring complex coupling relationships within and between market behaviors. We discuss this in detail hereafter.

6.2.2 Problem Formalization

This chapter aims to forecast stock market trends based on hierarchical coupled market behaviors. The problem can be formalized as follows: the coupling function $f(\cdot)$ here is used to capture the complex hierarchical coupling relationships between different financial markets and countries, and the objective function $g(\cdot)$ is built to forecast the possibilities of two trends (upward and downward) for stock market on the following trading time, namely $g^S(up)$ represents the possibility of an upward trend in a stock market while $g^S(down)$ represents the possibility of a downward trend. If at time t,

$$g^S_{t+1}(up) \mid f_t(\cdot) \geq g^S_{t+1}(down) \mid f_t(\cdot), \tag{6.1}$$

then time $t+1$ is a upward trend, otherwise a downward trend.

In order to do the forecasting, our key task now is to build a proper model to determine the specific coupling function $f(\cdot)$ and the corresponding objective function $g(\cdot)$. Below, HCBA is introduced to understand the hierarchical coupled cross-market behaviors, some definitions are given, and then MCHMM is explored to capture the two-layer complex coupled relationships between multiple markets in different countries.

6.3 Modeling Framework

6.3.1 Hierarchical Cross-market Behavior Analysis

Coupled cross-market behaviors refer to the activities from multiple markets with inter- and intra-relations between markets, while *Hierarchical cross-market behaviors* represent the two-layer coupled cross-market behaviors, Layer-1 is intra-country market coupled behavior, namely the coupled behaviors between markets in the same country and Layer-2 is inter-country market coupling which represents the coupled behaviors among different countries. Suppose there are K countries $\{C_1, C_2, \ldots, C_K\}$, an country C_k owns I markets $\{M_1, M_2, \ldots, M_I\}$, a market M_i undertakes J market behaviors

$\{MB_{i1}, MB_{i2}, \ldots, MB_{iJ}\}$, market M_i's j^{th} behavior MB_{ij} is a q-variable vector, its variable $[p_{ij}]_q$ reflects the q^{th} behavior property. Then a Market Behavior Feature Matrix with country C_k $FM(MB^{C_k})$ is defined as follows:

$$FM(MB^{C_k}) = \begin{pmatrix} MB_{11}^{C_k} & MB_{12}^{C_k} & \ldots & MB_{1J_i}^{C_k} \\ MB_{21}^{C_k} & MB_{22}^{C_k} & \ldots & MB_{2J_i}^{C_k} \\ \vdots & \vdots & \ddots & \vdots \\ MB_{I1}^{C_k} & MB_{I2}^{C_k} & \ldots & MB_{IJ}^{C_k} \end{pmatrix}$$

Hence, Layer-1 coupling is the relationships within one Market Behavior Feature Matrix, including the relationships within one row and the couplings embodied through the columns. Layer-2 coupling is the coupled relationships between different Market Behavior Feature Matrixes.

Definition 6.1 Hierarchical market behaviors *(HMB) refer to two-layer coupled market behaviors. Level 1* MB^1 *is the intra-country market coupling, which represents the coupling relationships between different markets behaviors* $MB_{i_1j_1}^{C_1}$ *and* $MB_{i_2j_2}^{C_2}$ *within one country under* $f(\theta(\cdot))$ *, where* $(C_1 = C_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J_i) \wedge (1 \leq C_1, C_2 \leq K)$

$$MB^1 = MB_{i_1j_1}^{C_1} \odot MB_{i_2j_2}^{C_2} | f(\theta(\cdot)) \tag{6.2}$$

where $\odot$ *means the interactions of* $MB_{i_1j_1}^{C_1}$ *and* $MB_{i_2j_2}^{C_2}$.

Level 2 MB^2 *is inter-country market coupling, which represents market behaviors* $MB_{i_1j_1}^{C_1}$ *and* $MB_{i_2j_2}^{C_2}$ *that are coupled in terms of relationships* $f(\eta(\cdot))$*, where* $(i_1 = i_2) \vee (C_1 \neq C_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J_i) \wedge (1 \leq C_1, C_2 \leq K)$

$$MB^2 = MB_{i_1j_1}^{C_1} \odot MB_{i_2j_2}^{C_2} | f(\eta(\cdot)) \tag{6.3}$$

Then

$$HMB ::= (MB^1, MB^2) \mid f(\theta(\cdot), \eta(\cdot)) \tag{6.4}$$

Here $f(\theta(\cdot), \eta(\cdot))$ *is the coupling function denoting the corresponding relationships between different* MB*.* $\theta(\cdot)$ *is the intra-country market coupling* *function, capturing the coupled relationships between markets in the same country.* $\eta(\cdot)$ *is the inter-country market coupling function, representing the coupled relationships between the same markets in various countries.*

Definition 6.2 Hierarchical Cross-Market Behavior Analysis *(HCBA) is to build the objective function* $g(\cdot)$ *under the condition that market behaviors between different markets and various countries are coupled with each other by coupling function* $f(\cdot)$*, and satisfy the following conditions:*

$$f(\cdot) ::= f(\theta(\cdot), \eta(\cdot)) \tag{6.5}$$

$$g(\cdot)|f(\cdot) \geq g_0|f_0 \tag{6.6}$$

where $\theta(\cdot)$ *is the Layer-1 coupling, while* $\eta(\cdot)$ *captures Layer-2 coupling relationships.*

The above discussion gives the basic concept of HCBA, and the vector-based market behavior representation shows us a road map from understanding to modeling HCBA in real financial markets. In the following we develop a MCHMM to model the HCBA and the specific mapping process will be introduced in Section 6.3.

6.3.2 Multi-layer Coupled Hidden Markov Model

The standard CHMM is briefly introduced in Section 2.3.1, and more details can be found in (Zhong & Ghosh 2001). It is often used to model multiple processes with coupling relationships, but it can not handle the complex hierarchical coupled relationships between markets in different countries, as mentioned above. Here we propose a Multi-layer CHMM (MCHMM) architecture to fit the HCBA framework for solving problems such as capturing complex couplings between markets within one country and between countries.

As shown in Figure 6.3 with two countries (C_1, C_2) as an example, each country owns two financial markets M_1, M_2. There exists two-layer coupled behaviors, Layer-1 is intra-country market coupling from the market aspect, where the state of one market in a country depends on the state of the markets in the same country (e.g. the state of M_1 in C_1 at time $t+1$ $Z_{t+1}^{C_1M_1}$ depends on $Z_t^{C_1M_1}$ and $Z_t^{C_1M_2}$). Layer-2 is inter-country market coupling from the country perspective, namely the state of one market in one country also relies on the state of the same markets in other countries (e.g. the state $Z_{t+1}^{C_1M_1}$ relies on $Z_t^{C_2M_1}$).

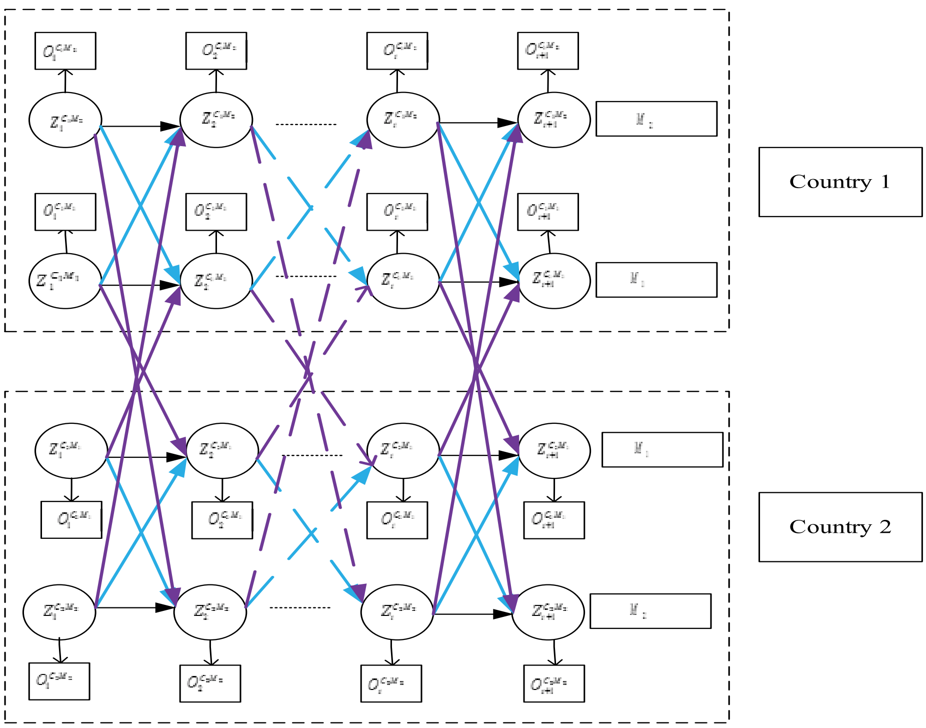


Figure 6.3: An example of MCHMM

The above discussion suggests the urgent need to build a MCHMM to model the complex hierarchical coupled interactions among different markets in various countries. Suppose there are K countries $\{C_1, C_2, \dots, C_K\}$, a country C_k owns I markets $\{M_1, M_2, \dots, M_I\}$, and the elements of a MCHMM are as follows:

Prior probability of initial state $\pi = \{\pi_j^{(C_kM_i)}\}$

$\pi_j^{(C_kM_i)} = p(z_1^{(C_kM_i)} = Z_j)$, s.t. $\sum_{j=1}^{H^{(C_kM_i)}} \pi_j^{(C_kM_i)} = 1$
State transition probability matrix $A = \{a_{j'j}^{(C_{k'}M_{i'},C_kM_i)}\}$
$a_{j'j}^{(C_{k'}M_{i'},C_kM_i)} = p(z_{t+1}^{(C_kM_i)} = Z_j | z_t^{(C_{k'}M_{i'})} = Z_{j'})$,
s.t. $\sum_{j=1}^{H^{(C_kM_i)}} a_{j'j}^{(C_{k'}M_{i'},C_kM_i)} = 1$
Observation probability matrix $B = \{b_j^{(C_kM_i)}(v)\}$
$b_j^{(C_kM_i)}(v) = p(o_t^{(C_kM_i)} = X_v | z_t = Z_j)$,
s.t. $\sum_{v=1}^{V} b_j^{(C_kM_i)}(v) = 1$
Coupling weight $CR = \{CR_1, CR_2\}$
CR denotes the weight of mixture where
$CR_1 = \{r_{C_kM_{i'},C_kM_i}\}$ is the intra-country coupling weight,
$CR_2 = \{r_{C_{k'}M_i,C_kM_i}\}$ is the inter-country coupling weight,
s.t. $\sum_{r_{i'} \in CR_1} r_{i'} + \sum_{r_{k'} \in CR_2} r_{k'} = 1$

H is the number of states of the Markov chains, $\{Z_1, Z_2, \ldots, Z_H\}$ is a set of hidden states, where z_t is the hidden state at time t. V is the number of observation symbols, $\{X_1, X_2, \ldots, X_V\}$ is a set of observation symbols, $O = \{O_1, O_2, \ldots, O_T\}$ is an observation sequence, o_t is the observation at time t.

Then the state joint transition probability is

$$P(Z_{t+1}^{C_kM_i} | Z_t^{C_{k'}M_i}, Z_t^{C_kM_{i'}}) = \sum_{r_{i'} \in CR_1} r_{i'} P(Z_{t+1}^{C_kM_i} | Z_t^{C_kM_{i'}}) + \sum_{r_{k'} \in CR_2} r_{k'} P(Z_{t+1}^{C_kM_i} | Z_t^{C_{k'}M_i}) \quad (6.7)$$

Based on the above discussion, similar to standard CHMM, below we refer to the complete set of parameters of a MCHMM as $\Omega = \{A, B, CR, \pi\}$.

6.3.3 Parameter Estimation

Here we can derive the $P(O|\Omega)$ by extending the forward-backward procedure in (Zhong & Ghosh 2001). The corresponding forward variable is calculated as follows:

$$\alpha_t^{(C_k M_i)}(j) = (\sum_{r_{i'} \in CR_1} r_{i'} \sum_{j'} (\alpha_{t-1}^{(C_k M_{i'})} \cdot a_{j'j}^{(C_k M_{i'}, C_k M_i)}) + \sum_{r_{k'} \in CR_2} r_{k'} \sum_{j'} (\alpha_{t-1}^{(C_{k'} M_i)} \cdot a_{j'j}^{(C_{k'} M_i, C_k M_i)}) b_j^{(C_k M_i)}(o_t) \tag{6.8}$$

Then, $P(O|\Omega)$ can be represented by forward variables:

$$P(O|\Omega) = \prod_{k,i} (\sum_j \alpha_T^{(C_k M_i)}(j)) \tag{6.9}$$

From the above we can see that the corresponding parameters $\omega \in \Omega$ satisfy equality constraints, i.e. the sum equals to one. This leads us to learn the parameters by constrained optimization techniques. Here we solve it by the Lagrange multiplier method, following the discussion in (Levinson, Rabiner & Sondhi 1983). It can be verified that $P(O|\Omega)$ is locally maximized when

$$\omega = \frac{\omega \partial P(O|\Omega)/\partial\omega}{\sum \omega \partial P(O|\Omega)/\partial\omega} \tag{6.10}$$

where $\partial P(O|\Omega)/\partial\omega = \sum_{k,i} (\frac{P(O|\Omega)}{\partial \alpha_T^{(C_k M_i)}(j)} \sum_j \frac{\partial \alpha_T^{(C_k M_i)}(j)}{\partial\omega})$ is derived similar to (Zhong & Ghosh 2001).

When each $\omega \in \Omega$ is estimated by Equation (6.10), we finish one iteration of optimization. $P(O|\Omega)$ is guaranteed to converge to some local maximum by iteratively updating w.r.t. Ω (cf. Theorem 2 in (Zhong & Ghosh 2001)).

6.4 Forecasting Methodology

This section describes the methodology to forecast the US stock market trends: data preprocessing, modeling the cross-market behaviors using M-CHMM, and forecast market trends.

6.4.1 Data Preprocessing

To better fit the financial time series data into the MCHMM, here we mainly focus on the following two parts.

A. Index Selection

As mentioned above, in this chapter there are 13 major countries and each country has two financial markets (stock market and currency market). In the MCHMM, we use one Markov chain to represent a market, and here we select two countries which have higher correlations with the stock market in the target country (US stock market) to build Layer-2 coupled relationships (inter-country coupling). This is because our focus is on the coupling relationships among various countries, hence higher correlation with the target market encloses a stronger discriminative power.

Definition 6.3 Country Index Correlation *(CIC) Suppose there are* K *countries* $\{C_1, C_2, \ldots, C_K\}$*, an country* C_k *owns* I *markets* $\{M_1, M_2, \ldots, M_I\}$*,* C_kM_i *is the target market in target country,* $CIC_{C_{k'}M_i, C_kM_i}$ *refers to the correlations of the target market indexes in other countries* $CI_{C_{k'}M_i}$ *with* $CI_{C_kM_i}$*, where* $(k' \neq k) \wedge (1 \leq k, k' \leq K) \wedge (1 \leq i \leq I)$*.*

$$CIC_{C_{k'}M_i, C_kM_i} = \rho_{DCCA}(CI_{C_{k'}M_i}, CI_{C_kM_i}) \tag{6.11}$$

where ρ_{DCCA} *is the cross-correlation coefficient indicated in Section 6.2.1.*

B. Data Decoding

The data we choose for each index is the weekly closing price. Since the data may show various data types in different financial markets among various countries, this may lead to an extremely large set which would make the MCHMM exceedingly complex. Therefore, to better fit the MCHMM used after, it is imperative to find a way to encode the observed closing price into a set of symbols suitable for the MCHMM. Many researchers choose to use the return as the symbols which can be calculated by $RI_t = \frac{PI_t - PI_{t-1}}{PI_{t-1}} * 100\%$, here RI_t and PI_t are, respectively, the return and closing price at time t.

In this chapter, we will use a discretization method to replace the return. This is because from the distribution perspective, the price change is not a uniform distribution, namely the returns which is a linear segment can not represent it correctly. For example, the stock price change limit is 10%, but ^DJI weekly changes are around [-2%, 2%] from Jan 2003 to Dec 2013, and very limited changes achieve around 10% (-10%), which means the group [-2%, 2%] owns more data and the length of the segment should not be linear. So it is not proper to choose the returns as the symbols.

The mapping of weekly returns into discrete symbols is similar to the quartile, which is a type of quantile. It is calculated as follows: we use four points that divide the data set into five groups: the first point is defined as the middle number between the smallest number and the median of the data set; the second point is the median of the data; the third point is the middle value between the median and the highest value of the data set; and the fourth point is the 0[4]. After this, each interval is associated with a different symbol, for example $\{-2, -1, 1, 2, 3\}$ (upward symbol $O^+ = \{1, 2, 3\}$ and downward symbol $O^- = \{-2, -1\}$), and every return belongs to a particular interval is associated to the same symbol.

6.4.2 MCHMM-based Hierarchical Market Behavior Modeling

Here, we model the forecasting problem in terms of HCBA and the MCHMM, which can be divided into two mapping processes: one from HCBA to the MCHMM, and another from the MCHMM, to the specific stock market trend forecasting problem. The first mapping process is to map the HCBA-based cross-market behaviors to a MCHMM model which captures the hierarchical coupling relationships between markets within and between countries, and the second is to handle the forecasting problem with the produced MCHMM.

A. Mapping Process from HCBA to MCHMM

Following the general problem definition of cross-market behavior model by HCBA, here we model the couplings in cross-market in terms of Layer-1 and

[4]It is the break point of uptrend and downtrend.

Layer-2. We select two countries (C_1, C_2) which have the higher correlations with the target country C_u, and two markets (stock market and currency market) from each country. Correspondingly, there are two Markov chains built for each country, namely HMM-C_uM_s and HMM-C_uM_c enclosing the stock market sequences $(\Phi(C_uM_s))$ and currency market sequences $(\Phi(C_uM_c))$ in the target country respectively, HMM-C_1M_s, HMM-C_1M_c representing the two sequences in country C_1, and HMM-C_2M_s, HMM-C_2M_c capturing the two sequences in country C_2. The representation of coupled market behavior sequences in Equation (6.3) and (6.4) shows the possibility of using MCHMM to model HCBA. The mapping from HCBA to MCHMM works in the following way:

$$HCBA \rightarrow MCHMM\,modeling$$

$$f(\theta(\cdot), \eta(\cdot)) \rightarrow \Omega = \{A, B, CR, \pi\} \tag{6.12}$$

$$\Phi(\cdot)|transition \rightarrow A(Transitional\,probability) \tag{6.13}$$

$$\Phi(\cdot)|observation \rightarrow B(Observation\,probability) \tag{6.14}$$

$$r_{ij}|intra - country \rightarrow CR_1 \tag{6.15}$$

$$r_{ij}|inter - country \rightarrow CR_2 \tag{6.16}$$

$$\Phi(\cdot)|prior \rightarrow \pi(Prior\,Probability) \tag{6.17}$$

$$g_{t+1}^{C_uM_s}(\cdot) \mid f_t(\cdot) \rightarrow P(g_{t+1}^{C_uM_s} \mid \Omega) \tag{6.18}$$

where $i, j \in \{C_uM_s, C_uM_c, C_1M_s, C_1M_c, C_2M_s, C_2M_c\}$, r_{ij} represents the coupling coefficients between multiple markets in various countries.

B. Mapping Process from MCHMM to US Stock Market Forecasting

With the MCHMM model built on top of the HCBA framework for cross-market behavior analysis, here we explain how to conduct forecasting using the MCHMM. We build a mapping process from MCHMM to the specific stock market forecasting. In the MCHMM, several HMMs are built for the hierarchical cross-markets. Suppose there are H hidden states in an HMM, which are denoted as $Z = Z_1, Z_2, \ldots, Z_H$, where Z_j is an individual state, and the state at time t is denoted as z_t. $O = \{O_1, O_2, \ldots, O_T\}$ is an observation sequence, o_t is the observation at time t. Below, we discuss the specific mapping relationships.

$$MCHMM \rightarrow Stock\ Market\ Forecasting\ Mapping$$

$$A \rightarrow P(z_{t+1}^{cm} = Z_j \mid z_t^{c'm'} = Z_{j'}) \tag{6.19}$$

$$B \rightarrow P(o_t^{cm} \mid z_t^{cm} = Z_j) \tag{6.20}$$

$$CR_1 \rightarrow r(cm, cm') \tag{6.21}$$

$$CR_2 \rightarrow r(cm, c'm) \tag{6.22}$$

$$\pi \rightarrow \pi^{cm} \tag{6.23}$$

$$P(g_{t+1}^{C_uM_s} \mid \Omega) \rightarrow P(o_{t+1}^{C_uM_s} \mid O_{1:t}, \Omega) \tag{6.24}$$

where $(c'm', cm \in \{C_uM_s, C_uM_c, C_1M_s, C_1M_c, C_2M_s, C_2M_c\}) \wedge (c \neq c') \wedge (m \neq m'), 1 \leq i, j \leq H$. $r(cm, c'm)$ represents the correlations between the same markets in different countries while $r(cm, cm')$ represents correlations between different markets in same country. $P(o_{t+1}^{C_uM_s} \mid O_{1:t}, \Omega)$ is to forecast stock market movement at time $t+1$ based on the observations from all market sequences in the time interval $[1, t]$ $(O_{1:t})$. Then if the upward trend probability at time $t+1$ $(P(o_{t+1}^{C_uM_s} \in O^+))$ is larger than the downward trend probability at time t $(P(o_{t+1}^{C_uM_s} \in O^-))$, time $t+1$ is an upward trend, otherwise the downward trend.

6.4.3 Stock Market Forecasting Process

The above section explains how to convert the HCBA framework to stock market trend forecasting by implementing a MCHMM. Below we discuss the specific process for the forecasting, namely how to calculate the $P(o_{t+1}^{C_uM_s}) \mid O_{1:t}, \Omega)$.

Figure 6.4 illustrates the general framework of the proposed forecasting process. At each time interval $[1, t]$ (the length here is T) , the first step is to train the MCHMM using the most recent T observations ($O_{1:t}$) in the six market sequences mentioned above, and then we can obtain the current model $\Omega = \{A, B, CR, \pi\}$ (Every T days, the parameters in the MCHMM will be reestimated using the most recent T observations). After this, based on the trained model Ω and the T observations $O_{1:t}$, the MCHMM gives the probability distribution for the following movements of the stock market in the target country at time $t+1$. Then we can obtain the stock market trend at time $t+1$ through comparing the probabilities of forecasted the upward trend $P(o_{t+1}^{C_u M_s} \in O^+)$ and the downward trend $P(o_{t+1}^{C_u M_s} \in O^-)$ at time $t+1$. w is the number of forecasting points, namely the number of sliding windows. The specific deviation steps are as follows[5]:

$$\begin{aligned}
P(o_{t+1}^{C_u M_s} \mid O_{1:t}) &= \sum_{z_{t+1}} p(o_{t+1}^{C_u M_s}, z_{t+1}^{C_u M_s} \mid O_{1:t}) \\
&= \sum_{z_{t+1}} p(o_{t+1}^{C_u M_s} \mid z_{t+1}^{C_u M_s}) p(z_{t+1}^{C_u M_s} \mid O_{1:t}) \\
&= \sum_{z_{t+1}} p(o_{t+1}^{C_u M_s} \mid z_{t+1}^{C_u M_s}) \sum_{\mathbf{z_t}} p(z_{t+1}^{C_u M_s}, \mathbf{z_t} \mid O_{1:t}) \\
&= \sum_{z_{t+1}} p(o_{t+1}^{C_u M_s} \mid z_{t+1}^{C_u M_s}) \sum_{\mathbf{z_t}} p(z_{t+1}^{C_u M_s} \mid \mathbf{z_t}) p(\mathbf{z_t} \mid O_{1:t}) \\
&= \sum_{z_{t+1}} p(o_{t+1}^{C_u M_s} \mid z_{t+1}^{C_u M_s}) \sum_{\mathbf{z_t}} p(z_{t+1}^{C_u M_s} \mid \mathbf{z_t}) \frac{p(\mathbf{z_t}, O_{1:t})}{p(O_{1:t})}
\end{aligned} \tag{6.25}$$

where $\mathbf{z_t}$ is the states at time t from all the market sequences mentioned above.

Equation (6.25) can be evaluated with the following steps: firstly, we will run a forward recursion, from which we will obtain $P(\mathbf{z_t}, O_{1:t}) = \prod_{k,i} \alpha^{C_k M_i}(\mathbf{z_t})$ and $P(O_{1:t}) = \prod_{k,i} \sum_{\mathbf{z_t}} \alpha^{C_k M_i}(\mathbf{z_t})$. After this, we compute the final summation over $\mathbf{z_t}$ and z_{t+1}. Then we will obtain the probabilities of upward and downward trends at time $t+1$. Once the observation at time $t+1$ is observed,

[5]Here we omit the Ω, rather than fixing it for simplicity. The conditional independence properties are used (Jordan 2002).

we can run the α recursion forward to the next step in order to predict the subsequent movement at time $t+2$, and this can carry forward indefinitely.

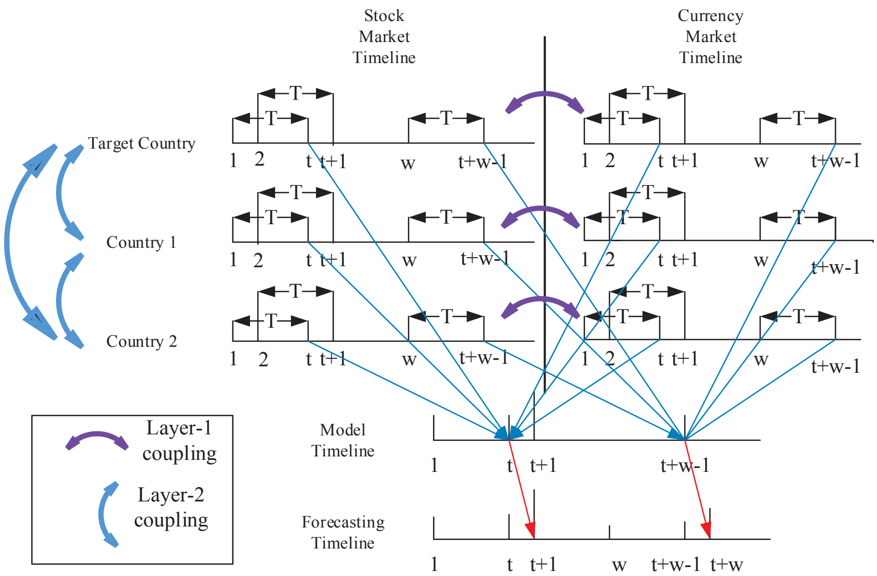


Figure 6.4: HCBA-based Forecasting Process

6.4.4 The Forecasting Algorithm

With the probabilities obtained from Equation (6.25), here we further determine the specific follow-up trend of the stock market through comparing the probabilities of the following two trends (upward and downward). The corresponding algorithm is described in Algorithm 6.1. The input is the observations of all market sequences $O^i(O^i = O_{i:t+i-1})$, and the time interval is T. It is a loop process to obtain w times of forecasting, namely the number of forecasting windows is w. We firstly train the model $\Omega^i(1 \leq i \leq w)$ according to the observations $O^i(O^i = O_{i:t+i-1})$, then we forecast the probabilities of two trends respectively. The output of the algorithm includes two sets: upward set UWS and downward set DWS, and the total objects in the two sets is w.

Algorithm 6.1 Stock Market Trend Forecasting via HCBA

Require: An observation set $\{O^1, O^2, \ldots, O^w\}$

```
for O^i in the observation set do
    Train the model Ω^i on the specific O^i;
    Compute the probability of the two trends given the model Ω^i and observations O^i at time t+i, respectively:
    P^{up}_{t+i} = P_{t+i}(o^{C_u M_s}_{t+1} ∈ O^+ | O^i, Ω^i) and P^{down}_{t+i} = P_{t+i}(o^{C_u M_s}_{t+1} ∈ O^- | O^i, Ω^i)
    if P^{up}_{t+i} ≥ P^{down}_{t+i} then
        trend at time t+i → UWS
    else
        trend at time t+i → DWS
    end if
end for
return An upward trend set UWS; A downward trend set DWS
```

6.5 Experiments

6.5.1 Experimental Settings

A. Data Set

In this section, we illustrate the use of the MCHMM for predicting US stock market movements based on capturing the hierarchical coupling relationships between the financial markets in various countries. Thus, the data set of interest is the historical prices of indexes in different financial markets. We choose two countries Netherlands and Germany which have the high correlations with the US based on Equation (6.11), and two types of markets: the stock market and currency market within these three countries for the case studies.

The data set used in this section includes weekly closing prices from Jan 2003 to Dec 2013, obtained from the Economic Research[6], and the prices are

[6]http://research.stlouisfed.org/

decoded into symbols based on Section 6.4.1. As indexes in different markets may appear on different trading days, we delete those days on which some market data is missing and only choose the days with trading data from all financial markets.

Table 6.2: MCHMM Elements Specification

Element		Element	
Country	3	Markov Chains	6
Market	2	Hidden State	10
Testing period	260 weeks	Forecasting Interval	39

Parameter Specification

- *Specification of the MCHMM Elements* The MCHMM elements are shown in Table 6.2. As mentioned in Section 6.2, there are three related countries and each country has two markets, so the number of Markov chains is six. In this chapter, the number of states is set equal to 10 based on the tests in the experiments. In addition, the size of the forecasting window, namely the forecasting interval T, was chosen which is neither too long (because of the uncertainty), nor too small (because of the high volatility of prices). So, according to the domain knowledge and several tests in the experiments, here we set T equal to 39 weeks.

- *MCHMM Initial Parameter Settings* Good starting values for parameters in the algorithm can help in speeding up the algorithm and ensuring promising results. Several possible kinds of initializations have been proposed. Using random starting values for the parameters and starting the algorithm from several different starting points and then selecting a better one is often used by researchers (Harshman & Lundy 1994). Here the initial parameter value of π, A and B follow the random selected method.

6.5.2 Comparative Methods

We compare the technical and business performance of our MCHMM-based approach with the following approaches:

- *ARIMA* We use ARIMA as the baseline method.
- *Logistic* We use this approach with indicators from different markets from various countries, and the parameters are calculated by MLE.
- *ANN* Here we will use ANN similar to (Hyup Roh 2007), with the cross-market data series.
- *CHMM* The different performance of CHMM and MCHMM will reflect the effect of hierarchical coupled analysis (the inter-country couplings). Here we use CHMM as a benchmark, and the specific steps refer to (Zhong & Ghosh 2001).
- *Buy and Hold* The buy and hold strategy is a long-term investment strategy based on the view that in the long run financial markets give a good rate of return despite periods of volatility or decline. In this chapter, the investor uses the strategy to buy the stock market index in the beginning and maintains this position during the whole trial.

6.5.3 Evaluation Metrics

We compare the performance of the MCHMM based approach and other approaches from technical and business perspectives.

A. Technical Perspective

- *Accuracy.* Accuracy=$\frac{TN+TP}{TP+FP+FN+TN}$
- *Precision.* Precision=$\frac{TP}{TP+FP}$. where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively. We treat the upward trend cases as the positive class here.

- *Type I error.* The percentage of the number of times with no upward forecasting when there is an upward trend against the times that there is an upward trend.

B. Business Perspective

Here, we analyze the returns obtained by an investor who uses the predictive outcomes of each approach to make trading decisions. The trading strategy adopted by the investor is as follows: if the approach forecasts an upward trend, the investor makes a buy decision in the stock market, otherwise, if there is a downward trend from the forecasting, a sell action is taken.

- *Rate of Return (ROR).* $ROR = \frac{Final\ Captical - Initial\ Capital}{Initial\ Capital}$.
- *Annualized Rate of Return (ARR).* It is the arithmetic mean of a series of rates of return.
  $ARR = \frac{Return\ in\ Period\ A + \cdots + Return\ in\ Period\ N}{Number\ of\ Periods}$.

6.5.4 Experimental Results

A. Technical Performance

Here, we compare the technical performance of our approach against the other four approaches on the testing period (Jan 2009 to Dec 2013). Accuracy, precision and type I error in the former part are calculated. The results are illustrated in Tables 6.3 and 6.4, and Figures 6.5 and 6.6.

Table 6.3: Technical Performance Comparison in Stock Market

Approach	Accuracy	Precision	Type I error
ARIMA	0.5476	0.6185	0.3815
Logistic	0.5794	0.6287	0.3713
ANN	0.5913	0.6450	0.3550
CHMM	0.6230	0.6852	0.3148
MCHMM	**0.6865**	**0.7159**	**0.2841**

Table 6.3 shows the performance of the five approaches over the whole testing period, while Table 6.4 gives the accuracy of the five approaches in each financial year. From the two tables we can see that our MCHMM-based approach has the best performance, both over the whole time period and yearly. For instance, the MCHMM has the highest accuracy increase of about 15% improvement over the ARIMA approach, and has around a 10% gain over the Logistic and ANN methods.

Table 6.4: Accuracy Comparison Yearly in Stock Market

Year	Accuracy				
	ARIMA	Logistic	ANN	CHMM	MCHMM
01/09-12/09	0.5192	0.6154	0.6346	0.5577	**0.7308**
01/10-12/10	0.6923	0.5962	0.5769	0.7115	**0.7692**
01/11-12/11	0.5000	0.5385	0.5000	0.6346	**0.6538**
01/12-12/12	0.4038	0.5577	**0.6538**	0.5962	0.5962
01/13-12/13	0.6222	0.6444	0.5111	0.6000	**0.6667**

Figures 6.5 and 6.6 show the technical performance, where the horizontal axis (P-Num) stands for the number of detected upward trends, namely the number of trading weeks with upward trends in the US stock market, and the vertical axis represents the values of technical measures. We can see that the MCHMM performs the best, followed by the CHMM. For example, precision improvement could be as high as about 20% against the ARIMA approach, and around 5% against the CHMM method when P-Num is bigger.

The results in the above tables and figures lead to the following conclusions: our MCHMM-based approach has the best performance. This clearly shows that the HCBA is a promising approach in stock market trend forecasting. The main reason being that it can capture the hierarchical coupled hidden relationships between different financial markets in various countries. Also, the MCHMM is demonstrated as a useful tool to undertake HCBA.

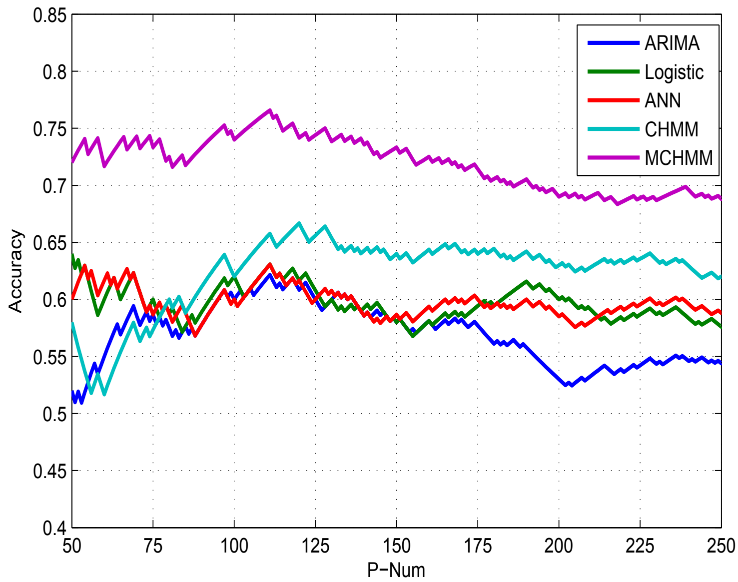


Figure 6.5: Accuracy of Various Approaches in Stock Market

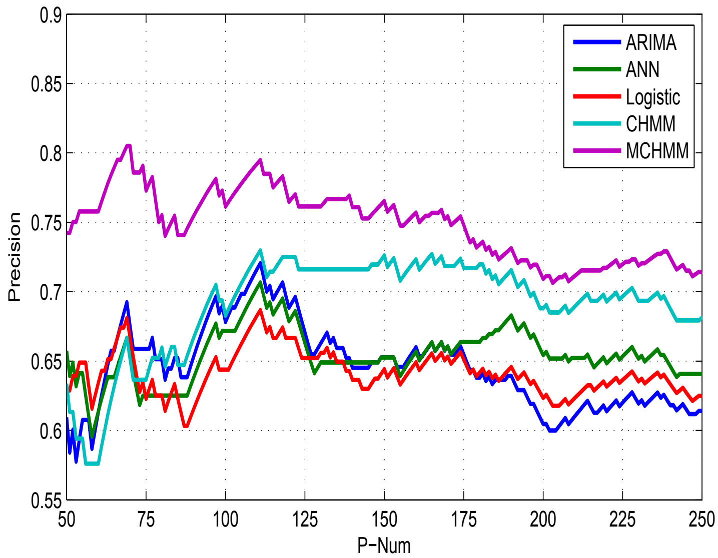


Figure 6.6: Precision of Various Approaches in Stock Market

B. Business Performance

Here, we compare the performance of our approach against the other five approaches from a business perspective. ROR, ARR and investor's wealth have been calculated in this part. The results of the business performance are reported in Tables 6.5 and 6.6 and Figure 6.7.

Table 6.5: ROR Comparison in Stock Market

Year	ROR				
	ARIMA	Logistic	ANN	CHMM	MCHMM
01/09-12/09	0.2450	0.1930	**0.3662**	0.2560	0.2930
01/10-12/10	0.1395	0.0228	0.0205	0.2427	**0.2649**
01/11-12/11	0.0225	-0.0103	0.0592	0.1280	**0.1910**
01/12-12/12	-0.0601	0.0506	**0.1262**	**0.1262**	0.067
01/13-12/13	0.1711	0.1845	0.1101	0.1524	**0.2343**

Tables 6.5 and 6.6 illustrate results of ROR and ARR, as ROR and ARR indicators are important for investors to validate the actionability of outcomes in real financial markets. The tables show that our approach performs much better than other approaches. For instance, our approach is with the highest ROR and has a gain of about 13% compared to the ARIMA method by 2010, and 8% over CHMM model by 2013. During the whole testing period, our MCHMM-based method has the biggest ARR, and higher by 8% than ANN and around 3% than the best performing CHMM.

Table 6.6: ARR Comparison in Stock Market

Year	ARR				
	ARIMA	Logistic	ANN	CHMM	MCHMM
01/09-12/09	0.2450	0.1930	**0.3662**	0.2560	0.2930
01/09-12/10	0.1923	0.1079	0.1934	0.2494	**0.2790**
01/09-12/11	0.1357	0.069	0.1486	0.2089	**0.2496**
01/09-12/12	0.0867	0.0640	0.1430	0.1882	**0.2040**
01/09-12/13	0.1036	0.0881	0.1364	0.1811	**0.2100**

Figure 6.7 shows an investor's wealth evolution by investing six trading strategies from Jan 2009 to Dec 2013. We have the following settings for the

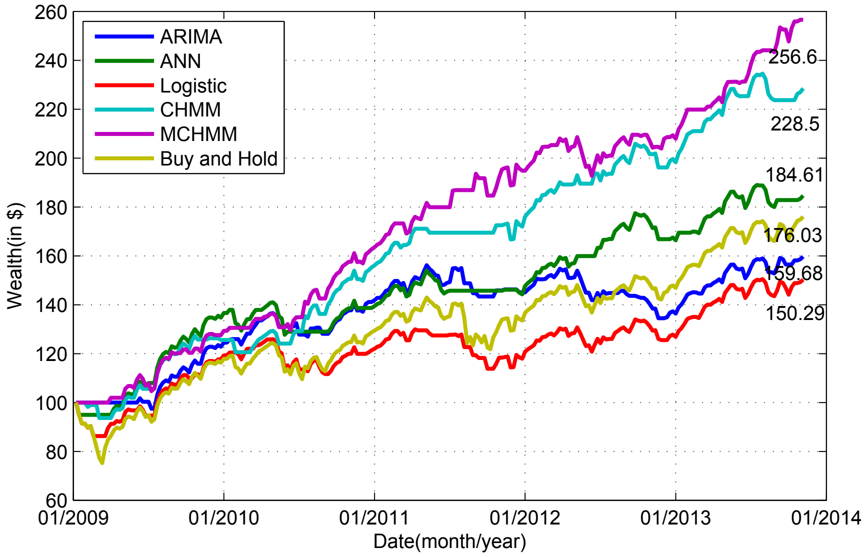


Figure 6.7: Investor's Wealth Comparison between Various Approaches

investment: (1) the initial capital investment is USD100; (2) no new capital will be added thereafter; (3) the investor buys and sells the index according to the trends forecasted by each approach (buy when there is an upward forecasting and sell while downward); (4) there are no transition fees. The following conclusions can be drawn from the figure: the MCHMM-based approach performs best, that is, an investor taking the recommendations from the approach can make profit at \$156.60, which represents a return of 156% in five years after the financial crisis period. MCHMM-based HCBA has been demonstrated to be a useful tool to help investors to make wiser trading decisions. In addition, our approach and the standard CHMM perform better than other five approaches, which verifies that hierarchical coupled relationships really exist among the financial markets in various countries.

6.6 Summary

Predicting market trends is very challenging especially under the fluctuation of financial markets due to financial recession and crisis. The analysis of hierarchical coupling relationships between relevant markets and multiple countries can contribute to the forecasting of market movements and resultant investment profitability. In this chapter, we propose a hierarchical coupled cross-market behavior analysis framework and a MCHMM to capture the complex hierarchical coupling relationships between various markets in different countries; the MCHMM-based forecasting model further predicts the movements of a market by considering the couplings with other markets across countries. The results show that the recommendations from the proposed method can gain better investment outcomes compared to those from ARIMA, Logistic, ANN, CHMM and the Buy and Hold strategy.

Chapter 7

Market Trend Forecasting via Coupled Temporal Belief Network

In this chapter, we investigate a deep architecture model to capture high-level coupled behaviors between different markets for market trends forecasting. The couplings involve interactions between homogeneous markets from various countries called intra-market coupling, interactions between heterogenous markets called inter-market coupling and interactions between current and past market behaviors called temporal coupling.

In this chapter, we design a coupled temporal deep belief network to hierarchically model the complex couplings. Multiple wings of conditional gaussian restricted boltzmann machines consist of the first layer, where each wing represents one type homogeneous market, to model the intra-market coupling and corresponding temporal coupling. The second layer is then built on the wings with coupled conditional restricted boltzmann machines, so as to conduct high-level inter-market coupling. Experimental results on data of stock and currency markets from three countries show that our approach outperforms other baselines, from both technical and business perspectives.

7.1 Background and Overview

The global financial crisis in 2008 and its contagion from the US to other regions and from one market to others show the importance of understanding the interactions across financial markets and the challenge of predicting future market movements. This is because financial markets are complex, evolutionary and non-linear dynamic systems; markets are no longer as independent as before due to globalization, there are explicit and implicit couplings between homogeneous and heterogeneous markets within and between countries. Accordingly, the price dynamics of a financial market cannot be simply informed by itself rather a systematic outcome of complex interactions across all related markets, as verified by the 2008 financial crisis (Longstaff 2010).

Figure 7.1 illustrates complex couplings across markets. In addition to other factors, the movement of US stock markets is affected by three major types of cross-market interactions: the *intra-market coupling*, referring to the interactions between homogeneous markets (e.g. UK stock market and Chinese stock market); the *inter-market coupling*, indicating the interactions between heterogeneous markets (e.g. US currency market and stock market); and the *temporal coupling*, describing the transitional influence across different time points in a market. Such couplings are embedded across different relevant markets and countries, which need to be considered in estimating the dynamics of a market.

However, it is very difficult to capture such couplings across financial markets. Let us explore the possible underlying challenges. Firstly, such complex interactions are driven by features that are not observable directly from market indexes (Chan et al. 2011); while we need to understand what such hidden factors (which may be abstract) are in order to find out the drivers of couplings. Secondly, the three types of couplings depicted in Figure 7.1 make it very difficult to build a model that is not too complex but expressive enough to capture the various interactions. Finally, these couplings behave in a highly non-linear and dynamic manner, which increases the difficulty to qualify them.

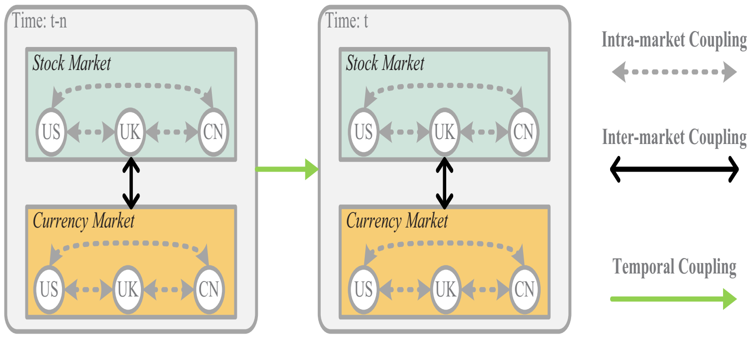


Figure 7.1: A Demonstration of Complex Couplings between Financial Markets

Modeling such couplings fundamentally challenges existing approaches for financial market forecasting, which can be roughly categorized into two groups: time series analysis represented by typical models including Logistic regression (Laitinen & Laitinen 2001), Autoregressive Integrated Moving Average (ARIMA) and Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) (Marcucci 2005) models, which use historical observations to infer future trend behaviors; and machine learning-based methods to forecast financial market movements, such as the applications of Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) (Hassan & Nath 2005). The main challenges lie in their deficiencies: linear time series models rely on the linear assumption of financial markets, which often exhibit nonlinear behaviors; more importantly, many models predict market trends directly based on observations, while ignoring the underlying complex couplings. As a result, the outcomes may be either biased or too sensitive to observations available. The machine learning-based models are shown more effective for capturing relationships within observations, especially non-linear interactions; however, limited work has been reported that it can jointly capture the above three types of interactions across markets.

To model such complex coupled relations, in this chapter, we propose to use the deep-learning approach to construct effective representation of cross-market couplings while maintains its reasonably size and expressiveness in capturing the three types of couplings. This is motivated by the recent theoretical results that show that the deep-architecture models (Bengio 2009) can be exponentially more efficient and expressive than shallow-structure ones. According to the couplings illustrated in Figure 7.1, we design a Coupled Temporal Deep Belief Network (CTDBN) to encode the *intra-market coupling*, *inter-market coupling* and *temporary coupling* within global financial markets. More specifically, in the first layer we employ Conditional Gaussian Restricted Boltzmann Machines (CGRBMs) to learn the abstract features to represent intra-market interactions and corresponding temporal dependence between homogenous markets. In the second layer, Coupled Conditional RBMs (CCRBMs) are built on the features learned from the first-layer models, so as to capture the high-level inter-market coupling between heterogeneous markets.

7.2 Preliminaries

We introduce some concepts used in this chapter and then formalize the complex interactions in financial markets. After this, we give a brief review on conditional RBM (CRBM) model which is the building block of our CTDBN model.

7.2.1 Problem Formalization

Suppose there are J countries, and each country owns I financial markets. m_{ij} represents the observations from market i in country j. In this chapter, we focus on representing three types of couplings (cf. Figure 7.1): *intra-market coupling*, *inter-market coupling* and *temporal coupling*. The corresponding definitions are as follows:

Definition 7.1 ***Intra-market Coupling***. *This is the interaction between homogeneous markets from all countries. Formally, the representation of*

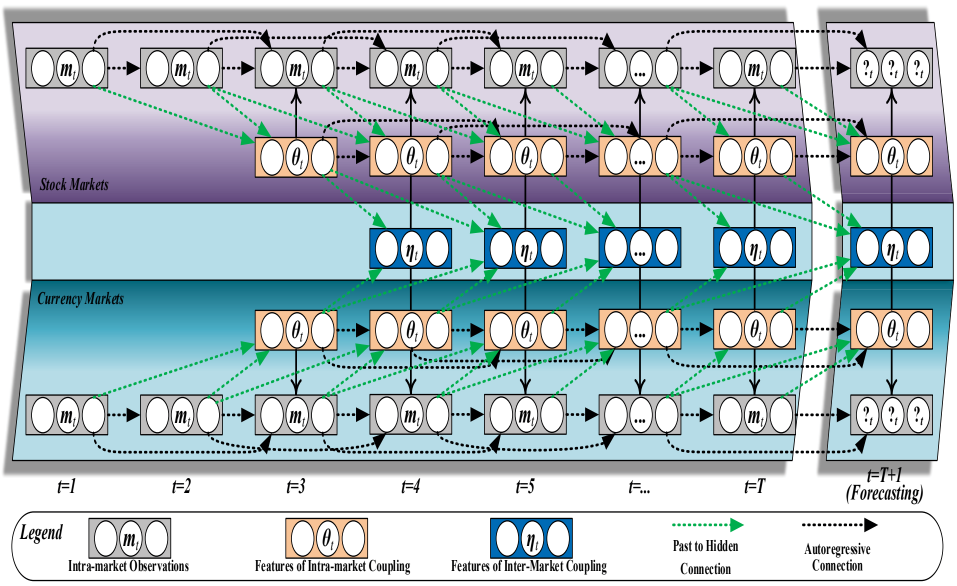


Figure 7.2: Modeling Framework of CTDBN. Here, the demonstration shows two heterogeneous financial markets, stock and currency. The first-layer are CGRBMs to model the intra-maket couplings while CCRBMs are built on the first layer to model inter-market couplings.

intra-market interaction w.r.t. the market i is given by:

$$\boldsymbol{\theta}_i = \{\otimes_{j=1}^{J}(\mathbf{m}_{ij})\} \tag{7.1}$$

where $\otimes$ denotes the coupled interactions among market i's observations over all countries.

Definition 7.2 ***Inter-market Coupling***. *This is the high-level interaction between heterogeneous financial markets, which is built on $\{\theta_i\}$. Formally, the representation of inter-market interaction is given by:*

$$\boldsymbol{\eta} = \{\circledast_{i=1}^{I}(\boldsymbol{\theta}_i)\} \tag{7.2}$$

where $\circledast$ denotes the couplings among all different markets.

Definition 7.3 ***Temporal Coupling***. *This denotes the influences from past information. The representation of n-order temporal coupling w.r.t. $\boldsymbol{\theta}_i$*

and $\boldsymbol{\eta}$ is given by:

$$\boldsymbol{\theta}_{i,t}|\{m_{ij,[t-n,t-1]}\}_{j=1}^{J} \tag{7.3}$$

$$\boldsymbol{\eta}_{t}|\{\boldsymbol{\theta}_{i,[t-n,t-1]}\}_{i=1}^{I} \tag{7.4}$$

which denotes the representation of coupled interaction at time t influenced by the past period from $t-n$ to $t-1$.

7.2.2 Conditional Restricted Boltzmann Machines

In order to model temporal coupling, we need to use CRBM (Taylor 2009) instead of RBM. The CRBM assign a probability to any joint setting of the visible units $\mathbf{v}$ and hidden units $\mathbf{h}$ conditional on $\mathbf{u}$ by

$$P(\mathbf{v},\mathbf{h} \mid \mathbf{u}) = exp(-E(\mathbf{v},\mathbf{h},\mathbf{u}))/Z \tag{7.5}$$

where Z is a normalization constant and $E(\mathbf{v},\mathbf{h},\mathbf{u})$ is an energy function:

$$E(\mathbf{v},\mathbf{h},\mathbf{u}) = -\mathbf{v}^{\mathbf{T}}\mathbf{W}\mathbf{h} - \mathbf{u}^{\mathbf{T}}\mathbf{A}\mathbf{v} - \mathbf{u}^{\mathbf{T}}\mathbf{B}\mathbf{h} - \mathbf{a}^{\mathbf{T}}\mathbf{v} - \mathbf{b}^{\mathbf{T}}\mathbf{h} \tag{7.6}$$

where $\mathbf{v} \in \{0,1\}^{D}$ is a vector of binary visible units, $\mathbf{h} \in \{0,1\}^{F}$ is a vector of binary hidden units and $\mathbf{u} \in \{0,1\}^{D}$ is a vector of binary visible units. $\mathbf{W} \in \mathbb{R}^{D\times F}$ encodes the interactions between $\mathbf{v}$ and $\mathbf{h}$, $\mathbf{A} \in \mathbb{R}^{D\times D}$ encodes the interactions between $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{B} \in \mathbb{R}^{D\times F}$ encodes the interactions between $\mathbf{u}$ and $\mathbf{h}$. $\mathbf{a} \in \mathbb{R}^{D}$ and $\mathbf{b} \in \mathbb{R}^{F}$ denote the biases of $\mathbf{v}$ and $\mathbf{h}$ separately. Hence, $\Omega = \{\mathbf{W},\mathbf{A},\mathbf{B},\mathbf{a},\mathbf{b}\}$ are the model parameters that need to learn.

The conditional distributions w.r.t. visible units and hidden units are factorial (Bengio et al. 2013), which can easily derived from Equation (7.5):

$$P(h_f = 1 \mid \mathbf{u},\mathbf{v}) = s(b_f + \mathbf{u}^{\mathbf{T}}\mathbf{B}_{:,f} + \mathbf{v}^{\mathbf{T}}\mathbf{W}_{:,f}) \tag{7.7}$$

$$P(v_d = 1 \mid \mathbf{v},\mathbf{u}) = s(a_d + \mathbf{u}^{\mathbf{T}}\mathbf{A}_{:,d} + \mathbf{W}_{d,:}\mathbf{h}) \tag{7.8}$$

where $s(x) = 1/1 + exp(-x)$ is the logistic function, $\mathbf{W}_{d,:}$ denotes the d_{th} row of $\mathbf{W}$ and $\mathbf{A}_{:,d}$ denotes the d_{th} column of $\mathbf{A}$. Such a notation will be used in the rest of this chapter.

Moreover, to model real-valued data (e.g. stock return), we need to employ a generalized CRBM with Gaussian visible units, so-called conditional Gaussian RBM (CGRBM). The corresponding energy function has the following form:

$$E(\mathbf{v},\mathbf{h},\mathbf{u}) = -\frac{\mathbf{v^T W h}}{\sigma} - \mathbf{u^T A v} - \mathbf{u^T B h} + \frac{(\mathbf{v}-\mathbf{a})^{\mathbf{T}}(\mathbf{v}-\mathbf{a})}{2\sigma^2} - \mathbf{b^T h} \tag{7.9}$$

where each visible unit $\mathbf{v_d} \in \mathbb{R}$, with the variance σ^2. Then the corresponding conditional distributions are given by:

$$P(h_f = 1 \mid \mathbf{v},\mathbf{u}) = s(b_f + \mathbf{u^T B}_{:,f} + \mathbf{v^T W}_{:,f}/\sigma) \tag{7.10}$$

$$P(v_d \mid \mathbf{v},\mathbf{u}) = \mathcal{N}(a_d + \mathbf{u^T A}_{:,d} + \sigma \mathbf{W}_{d,:}\mathbf{h}, \sigma^2) \tag{7.11}$$

In many applications, it is much easier to normalize each visible unit to zero mean and unit variance (Taylor & Hinton 2009), so that we can simply set $\sigma = 1$. In this chapter, we also preprocess our data following this way.

Parameter Estimation

Generally, the estimator is derived from a maximum likelihood learning procedure. Hence, we can minimize the following negative log-likelihood w.r.t. each parameter $\omega \in \Omega$:

$$-\frac{\partial logp(\mathbf{v})}{\partial \omega} = \mathbb{E}_{P(\mathbf{h}|\mathbf{v},\mathbf{u})}(\frac{\partial E(\mathbf{v},\mathbf{h},\mathbf{u})}{\partial \omega}) - \mathbb{E}_{P(\mathbf{v},\mathbf{h},\mathbf{u})}(\frac{\partial E(\mathbf{v},\mathbf{h},\mathbf{u})}{\partial \omega}) \tag{7.12}$$

The first term on the right hand, a.k.a. data-dependent expectation, is tractable but the second term, a.k.a. model-dependent expectation is intractable and must be approximated (Bengio et al. 2013). In practice, Contrastive Divergence (CD) (Hinton et al. 2006) can be used to approximate the expectation with a short $k - step$ (e.g. $k = 1$) Gibbs sampling using Equations (7.7), (7.8) and (7.10), (7.11), denoted as CD_k.

The stochastic gradient descent update using CD_k w.r.t. each parameter $\omega \in \Omega$ can be given by:

$$\omega \leftarrow \omega - \alpha(\frac{\partial E(\mathbf{v}^0, \mathbf{h}^0, \mathbf{u})}{\partial \omega} - \frac{\partial E(\mathbf{v}^k, \mathbf{h}^k, \mathbf{u})}{\partial \omega}) \tag{7.13}$$

where $\mathbf{v}^0$ are the visible data, $\mathbf{h}^0$ is sampled by Equation (7.7) or (7.10), $\mathbf{v}^k$ and $\mathbf{h}^k$ are sampled from the k-step Gibbs chain.

7.3 Modeling and Forecasting

In this section, we focus on modeling the coupled relations in the global financial markets as formalized by Definitions 7.1, 7.2 and 7.3. We design a CTDBN to hierarchically model such complex interactions. Figure 7.2 demonstrates a CTDBN that models the couplings between stock and currency markets in various countries. Similar to a DBN stacking RBMs layer by layer, our CTDBN consists of multiple wings of CRBMs in the first layer, where each wing models one type of homogeneous markets, e.g. stock markets in different countries. Obviously, the first-layer model is used to model the *intra-market coupling* and *temporal coupling*. The second layer model is built on these wings, i.e. heterogeneous markets, where the hidden units of the first layer serve as the visible units of the second layer CRBMs. Hence, the second layer model is used to model the *inter-market coupling* and *temporal coupling*. Thus, our CTDBN is eligible to represent all above defined coupled relations by such a deep structure.

7.3.1 Representation of Intra-market Coupling

Given a vector of observations m_i of the financial market i within a period $t-n$ to t (e.g. $n = 2$ represents three week stock market indexes) from J countries. Each element m_{ijt} denotes the observation of market i in country j at time t. Here, we employ CGRBMs to model the representation of *intra-market coupling* between the given homogeneous markets, as illustrated in

Figure 7.3. Note that we omit the subscript i for concise in following, when we focus on discussing the coupling in a specific financial market i.

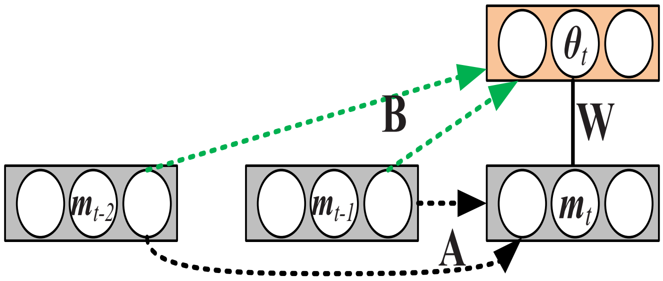


Figure 7.3: A CGRBM to Model Intra-market Coupling at Time t

As shown in Figure 7.3, a vector of Gaussian units is used to represent the current market observations $\mathbf{m}_t$. Moreover, to model the n-order *temporal coupling*, the past market observations $\mathbf{m}_{t-1}, \mathbf{m}_{t-2}, \ldots, \mathbf{m}_{t-n}$ are used to serve as the conditionals. Therefore, we use $\mathbf{A}_l$ for modeling the weights of autoregressive connections from $\mathbf{m}_{t-l}$ to $\mathbf{m}_t$. $\boldsymbol{\theta}_t \in \{0,1\}^F$ is a vector of hidden units serving as the abstract representation of *intra-market coupling*. In addition, the weights of connections from the past observations $\mathbf{m}_{t-l}$ to $\boldsymbol{\theta}_t$ are denoted as $\mathbf{B}_l$. Now, let $\mathbf{m}_{<t} = [\mathbf{m}_{t-1}, \mathbf{m}_{t-2}, \ldots, \mathbf{m}_{t-n}]$ denote a stacked history vector, and correspondingly the stacked weight matrices are denoted as $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_n]$ and $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_n]$. Therefore, the energy function of this CGRBM can be given as follows (we assume the variance equal to 1), according to Equation (7.9).

$$E(\mathbf{m}_t, \boldsymbol{\theta}_t, \mathbf{m}_{<t}) = \frac{(\mathbf{m}_t - \mathbf{a})^T(\mathbf{m}_t - \mathbf{a})}{2} - \mathbf{m}_t^T\boldsymbol{\theta}_t - \mathbf{m}_{<t}^T\mathbf{A}\mathbf{m}_t - \mathbf{m}_{<t}^T\mathbf{B}\boldsymbol{\theta}_t - \mathbf{b}^T\boldsymbol{\theta}_t \tag{7.14}$$

Then the conditional distributions can be immediately obtained according to Equations (7.10) and (7.11).

$$P(\theta_{ft} = 1 \mid \mathbf{m}_t, \mathbf{m}_{<t}) = s(b_f + \mathbf{m}_{<t}^T\mathbf{B}_{:,f} + \mathbf{m}_t^T\mathbf{W}_{:,f}) \tag{7.15}$$

$$P(m_{jt} \mid \boldsymbol{\theta}_t, \mathbf{m}_{<t}) = \mathcal{N}(a_j + +\mathbf{m}_{<t}^T\mathbf{A}_{:,j} + \mathbf{W}_{i,:}\boldsymbol{\theta}_t, 1) \tag{7.16}$$

A. Formalism Mapping

We easily find that the *intra-market coupling* operator $\otimes$ in Equation (7.1) is implemented by encoding $\{m_{jt}\}_{j=1}^{J}$ with the parameter $\mathbf{W}$, and the *temporal coupling* is encoded with the parameter $\mathbf{B}$. Therefore, $\boldsymbol{\theta}_t$ serves as features of *intra-market coupling* and associated *temporal coupling*.

B. Parameter Learning

Given Equations (7.14), (7.15) and (7.16), the stochastic gradient update equations using CD_k can be obtained according to Equation (7.13).

$$\mathbf{W} \leftarrow \mathbf{W} - \alpha(\mathbf{m}_t^{(0)}\boldsymbol{\theta}_t^{(0)T} - \mathbf{m}_t^{(k)}\boldsymbol{\theta}_t^{(k)T}) \tag{7.17}$$

$$\mathbf{A} \leftarrow \mathbf{A} - \alpha(\mathbf{m}_{<t}^{(0)}\mathbf{m}_t^{(0)T} - \mathbf{m}_{<t}^{(k)}\mathbf{m}_t^{(k)T}) \tag{7.18}$$

$$\mathbf{B} \leftarrow \mathbf{B} - \alpha(\mathbf{m}_{<t}^{(0)}\boldsymbol{\theta}_t^{(0)T} - \mathbf{m}_{<t}^{(k)}\boldsymbol{\theta}_t^{(k)T}) \tag{7.19}$$

$$\mathbf{a} \leftarrow \mathbf{a} - \alpha(\mathbf{m}_t^{(0)} - \mathbf{m}_t^{(k)}) \tag{7.20}$$

$$\mathbf{b} \leftarrow \mathbf{b} - \alpha(\boldsymbol{\theta}_t^{(0)} - \boldsymbol{\theta}_t^{(k)}) \tag{7.21}$$

C. Complexity Analysis

The whole time complexity of computation is divided into two major sections: energy compute (Equation (7.14)) and weight update (Equations (7.17)-(7.21)). The complexity of energy compute is $O(J \times F)$, and the complexity of weight update is $O(J \times F)$. Therefore, the total complexity is $O(J \times F)$.

7.3.2 Representation of Inter-market Coupling

As shown in Figure 7.1, *inter-market coupling* describes the interactions between heterogeneous markets. In fact, it can be viewed as a higher level relation that jointly models the coupling of all *intra-market couplings* as given by the Definition 7.2. Therefore, to model the representation of such a high-level *inter-market coupling*, we can build the second layer model on the first layer CGRBMs which hidden units, i.e. the representation of *intra-market coupling*, serve as the visible units of the second layer CRBM. In particular, the second layer CRBM couples all heterogeneous markets, so we call it Coupled CRBM (CCRBM).

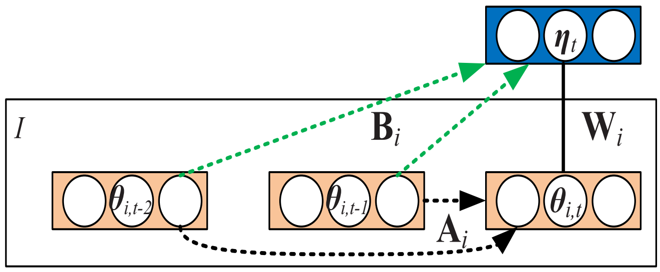


Figure 7.4: A CCRBM to Model Inter-market Coupling at Time t

Figure 7.4 illustrates the graphical model of the second layer CCRBM, where the plate notation is used to repeatedly represent the feature vector $\boldsymbol{\theta}_{i,t}$ learned from first layer CGRBMs w.r.t. each heterogeneous market i. $\boldsymbol{\theta}_t \in \{0,1\}^H$ is a vector of hidden units which serve as the features to represent *intra-market coupling*. $\boldsymbol{\eta}_t \in \{0,1\}^C$ is a vector of hidden units which serve as the features to represent *inter-market coupling*. Similar to the notation of first layer CGRBM, let $\boldsymbol{\theta}_{i,<t} = [\boldsymbol{\theta}_{i,t-1}, \boldsymbol{\theta}_{i,t-2}, \ldots, \boldsymbol{\theta}_{i,t-n}]$ denotes a stacked history vector w.r.t. market i. and $\mathbf{A}_i$, $\mathbf{B}_i$ are the stacked weight matrices associated with $\boldsymbol{\theta}_{i,<t}$ for modeling the *temporal couplings*. Then, we can write the energy function of this CCRBM as follows:

$$E(\{\boldsymbol{\theta}_{i,t}\}, \boldsymbol{\eta}_t, \{\boldsymbol{\theta}_{i,<t}\}) = -\mathbf{b}^T\boldsymbol{\eta}_t - \sum_{i=1}^{I} \mathbf{a}_i^T\boldsymbol{\theta}_{i,t} - \sum_{i=1}^{I} \boldsymbol{\theta}_{i,t}^T\mathbf{W}_i\boldsymbol{\eta}_t - \sum_{i=1}^{I} \boldsymbol{\theta}_{i,<t}^T\mathbf{A}_i\boldsymbol{\theta}_{i,t} - \sum_{i=1}^{I} \boldsymbol{\theta}_{i,<t}^T\mathbf{B}_i\boldsymbol{\eta}_t \tag{7.22}$$

According to the energy function, we can respectively obtain the conditional distribution w.r.t. each inter-coupled feature η_{lt}, and each intra-coupled feature θ_{rt}^j.

$$P(\theta_{ift} = 1 \mid \boldsymbol{\eta}_{ht}, \{\boldsymbol{\theta}_{i,<t}\}) = s(\mathbf{a}_{if} + \boldsymbol{\theta}_{i,<t}^T(\mathbf{A}_i)_{:,f} + (\mathbf{W}_i)_{f,:}\boldsymbol{\eta}_t) \quad (7.23)$$

$$P(\eta_{ht} = 1 \mid \{\boldsymbol{\theta}_{i,t}\}, \{\boldsymbol{\theta}_{i,<t}\}) = s(\mathbf{b}_h + \sum_{i=1}^{I} \boldsymbol{\theta}_{i,t}^T(\mathbf{W}_i)_{:,h} + \sum_{i=1}^{I} \boldsymbol{\theta}_{i,<t}^T(\mathbf{B}_i)_{:,h}) \quad (7.24)$$

A. Formalism Mapping

We can find that the coupling encoding operator $\circledast$ given in Equation (7.2) is implemented by term $\sum_{i=1}^{I} \boldsymbol{\theta}_{i,t}^T(\mathbf{W}_i)_{:,h}$ in Equation (7.24), which encodes the interaction between *intra-market couplings* to features for representing *inter-market coupling*. In addition, the *temporal coupling* from each heterogeneous market i is encoded with the parameter $\mathbf{B}_i$. Therefore, $\boldsymbol{\eta}_{ht}$ serves as features of the *inter-market coupling* and associated *temporal coupling*.

B. Parameter Learning

The stochastic gradient update equations using CD_k can be obtained according to Equation (7.13).

$$\mathbf{W}_i \leftarrow \mathbf{W}_i - \alpha(\boldsymbol{\theta}_{i,t}^{(0)}\boldsymbol{\eta}_t^{(0)T} - \boldsymbol{\theta}_{i,t}^{(k)}\boldsymbol{\eta}_t^{(k)T}) \quad (7.25)$$

$$\mathbf{A}_i \leftarrow \mathbf{A}_i - \alpha(\boldsymbol{\theta}_{i,<t}^{(0)}\boldsymbol{\theta}_{i,t}^{(0)T} - \boldsymbol{\theta}_{i,<t}^{(k)}\boldsymbol{\theta}_{i,t}^{(k)T}) \quad (7.26)$$

$$\mathbf{B}_i \leftarrow \mathbf{B}_i - \alpha(\boldsymbol{\theta}_{i,<t}^{(0)}\boldsymbol{\eta}_t^{(0)T} - \boldsymbol{\theta}_{i,<t}^{(k)}\boldsymbol{\eta}_t^{(k)T}) \quad (7.27)$$

$$\mathbf{a}_i \leftarrow \mathbf{a}_i - \alpha(\boldsymbol{\theta}_{i,t}^{(0)} - \boldsymbol{\theta}_{i,t}^{(k)}) \quad (7.28)$$

$$\mathbf{b}_i \leftarrow \mathbf{b}_i - \alpha(\boldsymbol{\eta}_t^{(0)} - \boldsymbol{\eta}_t^{(k)}) \quad (7.29)$$

C. Complexity Analysis

The whole time complexity of computation is divided into two major sections: energy compute (Equation (7.22)) and weight update (Equations (7.25)-(7.29)). The complexity of energy compute is $O(H \times C)$, and the complexity of weight update is $O(H \times C)$. Therefore, the total complexity is $O(H \times C)$.

7.3.3 Forecasting Based on CTDBN

Our ultimate goal is to forecast the trends of financial markets derived from the underlying complex interactions. Our CTDBN is a generative model, where $\mathbf{m}_{i,t}$ is generated from the hidden units $\boldsymbol{\theta}_{i,t}$ as depicted by the Figure 7.2. Therefore, we firstly need to infer $\boldsymbol{\theta}_{i,T+1}$ so as to predict $\mathbf{m}_{i,T+1}$. Furthermore, $\boldsymbol{\theta}_{i,T+1}$ and $\boldsymbol{\eta}_{T+1}$ are jointly dependent, so we need to infer $\boldsymbol{\eta}_{T+1}$ as well. Since each layer is a n-order temporal model, we totally need $2n$ past observations, i.e. $[\mathbf{m}_{i,T+1-2n}, \ldots, \mathbf{m}_{i,T}]$.

In particular, we perform mean-field inference (Welling & Hinton 2002) to reconstruct $\boldsymbol{\theta}_{i,T+1}$ and $\boldsymbol{\eta}_{T+1}$ instead of a stochastic reconstruction to avoid sampling noise. The prediction steps are given as follows:

---

1. Estimate $\boldsymbol{\theta}_{i,t}$ for $1 \leq i \leq I$, $T+1-2n \leq t \leq T$.
Given the parameters $\{\mathbf{b}, \mathbf{B}, \mathbf{W}\}$ of the first layer CGRBM w.r.t market i, $\boldsymbol{\theta}_{i,t}$ is set to the mean of Equation (7.15)
$\boldsymbol{\theta}_{i,t} \sim s(\mathbf{b} + \mathbf{m}_{i,<t}^T \mathbf{B} + \mathbf{m}_{i,t}^T \mathbf{W})$
2. Initialize $\boldsymbol{\theta}_{i,T+1}$ for $1 \leq i \leq I$.
$\boldsymbol{\theta}_{i,T+1} \leftarrow \boldsymbol{\theta}_{i,T}$
3. Estimate $\boldsymbol{\theta}_{i,T+1}$ for $1 \leq i \leq I$ by K-iteration mean-field update on second layer CCRBM, c.f. Equations (7.23) and (7.24).
$\boldsymbol{\eta}_t \leftarrow s(\mathbf{b} + \sum_{i=1}^{I} \boldsymbol{\theta}_{i,T+1}^{(K)T} \mathbf{W}_i + \sum_{i=1}^{I} \boldsymbol{\theta}_{i,<t}^T \mathbf{B}_i)$
$\boldsymbol{\theta}_{i,T+1}^K \leftarrow s(\mathbf{a}_i + \boldsymbol{\theta}_{i,<T+1}^T \mathbf{A}_i + \mathbf{W}_i \boldsymbol{\eta}_t^K)$
4. Generate predicted observations $\mathbf{m}_{i,T+1}$ for $1 \leq i \leq I$.
Given the parameters $\{\mathbf{a}, \mathbf{A}, \mathbf{W}\}$ of the first layer CGRBM w.r.t market i, the prediction $\mathbf{m}_{i,T+1}$ is set to the mean of Equation (7.15)
$\mathbf{m}_{i,T+1} \leftarrow \mathbf{a} + \mathbf{m}_{<t}^T \mathbf{A} + \mathbf{W} \boldsymbol{\theta}_{i,t}$

So far we generate the forecasting of each market at time $t = T+1$. This procedure can carry forward indefinitely.

7.4 Experiments

7.4.1 The Data Sets

In this section, we illustrate the use of the CTDBN for predicting financial market movements based on capturing the complex interactions between different financial markets. Thus, the data set of interest is the historical prices of market indexes in various countries. In this chapter we choose five countries: USA and BRIC (Brazil, Russia, India and China), the reason choose BRIC here is the BRIC accounted for more than 25% of the world's total GDP according to the International Monetary Fund (IMF). Two types of markets: the stock market and currency market of each country is chosen [1], as shown in Table 7.1.

The data set used includes weekly closing prices from Jan 2007 to Dec 2013 [2] (364 observations), and the prices are decoded into returns by $RI_t = \frac{PI_t - PI_{t-1}}{PI_{t-1}} * 100\%$, here RI_t and PI_t are, respectively, the return and closing price at time t. As indexes in different markets may appear on different trading days, we delete those days on which some market data is missing and only choose the days with data from all financial markets.

Table 7.1: Trading Indexes from Five Countries

Country	Market	
	Stock Market	Currency Market
USA	^DJI	SDR/USD
Brazil	^BVSP	SDR/BRL
Russia	RTS.RS	SDR/RUB
India	^BSESN	SDR/INR
China	000001.SS	SDR/CNY

[1]Here we choose Special Drawing Right (SDR) as its numeraire is a potential claim on the freely usable currencies of IMF (Jang et al. 2011).

[2]http://research.stlouisfed.org/

7.4.2 Evaluation Metrics and Comparative Methods

Technical Perspective

- *Accuracy*. Accuracy=$\frac{TN+TP}{TP+FP+FN+TN}$, where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively. We treat the upward trend cases as the positive class here.

- *Precision*. Precision=$\frac{TP}{TP+FP}$.

- *Recall*. Recall=$\frac{TP}{TP+FN}$.

Business Perspective

We analyze the return gained by an investor who uses the predictive outcomes of each approach to trade the indexes. The trading strategy adopted by an investor is as follows: if an approach forecasts an upward trend, the investor takes a buy position in the index; otherwise, if there is a downward trend from the forecasting, a sell action is taken.

- *Annualized Rate of Return (ARR).*
  $ARR = \frac{Return\ in\ Period\ A + \cdots + Return\ in\ Period\ N}{Number\ of\ Periods}$.

Comparative Methods

To evaluate our approach, we take the following methods which are either typically used in financial markets or directly address market couplings:

- *ARIMA*: we use it as a baseline method.

- *Logistic*: We use this approach with indexes from the two markets in five countries, and the parameters can be obtained through MLE.

- *ANN* (Hyup Roh 2007): The model is trained with indexes from the two markets in five countries by using the back-propagation algorithm.

- *CHMM* (Zhong & Ghosh 2001): CHMM consists of multiple HMM chains, where each chain corresponds to model one type of financial market in a country.

- *CTDBN*: This is our deep learning approach, where the order n of the both CGRBM and CCRBM are set equal to 2 which yields good results in this experiment.

- *CGRBM*: This is a sub-model of CTDBN, which simply models the intra-market coupling by the first-layer CGRBMs without considering inter-market coupling.

7.4.3 Experimental Results

In this part, we present results in three countries. Therefore, our testing consists of two heterogeneous markets: stock and currency markets, w.r.t. USA, China and India. The testing data includes the financial crisis period (2007-2009) and a non-crisis period (2010-2013) (Here we split the data by years, and we use the last five years data as the training set before the testing year so as to learn the model parameters). This arrangement aims to disclose the model performance against different situations with interactions.

The results of Accuracy and ARR are reported in Table 7.2, where ARR is an important indicator for investors to validate the actionability of outcomes in real financial market. From both technique and business perspectives, the baseline method of ARIMA does not achieve a good performance, this is because ARIMA is built on stationary data (constant mean and variance), and pays no attention on the underlying complex hidden interactions between the different markets. For the similar reason, the Logistic and ANN approaches do not perform very well. Note that the ANN outperforms the Logistic approach, the main reason here is that Logistic approach is under a linear assumption, but the financial market, especially the hidden couplings, are not linear. The CHMM and CGRBM perform much better than Logistic

Table 7.2: Performance of Comparative Methods in US, China and India Markets

Model	Accuracy						ARR					
	Stock			Currency			Stock			Currency		
	US	China	India	US	China	India	US	China	India	US	China	India
ARIMA	0.5357	0.5071	0.5029	0.5471	0.5353	0.5214	-0.1356	0.0415	-0.0675	0.1479	-0.0116	0.0304
Logistic	0.5643	0.55	0.5196	0.6	0.6059	0.5386	0.0226	0.0796	0.0558	0.0269	0.0428	0.0645
ANN	0.6	0.6	0.5752	0.6235	0.6059	0.5747	0.1217	0.1486	0.0788	0.1332	0.1244	0.1032
CHMM	0.6533	0.6214	0.5852	0.6471	0.6353	0.5709	0.1934	0.1426	0.1132	0.1645	0.1498	0.1555
CGRBM	0.6357	0.6235	0.5898	0.6565	0.64	0.5932	0.1568	0.1526	0.1410	0.1758	0.1456	0.1660
CTDBN	**0.6729**	**0.6324**	**0.6258**	**0.6734**	**0.6535**	**0.6152**	**0.2073**	**0.1682**	**0.2261**	**0.1926**	**0.1792**	**0.1972**

and ANN, this is because they construct predictions on the hidden coupled features.

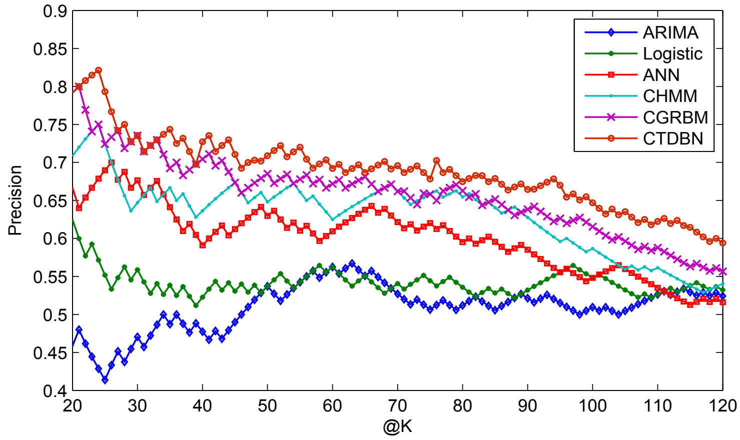
Our CTDBN outperforms all baselines regardless of technique or business perspective. This can be interpreted as follows: firstly, unlike those methods that predict market movements directly from the observations, CTDBN builds a deep architecture to learn the hidden features which removes the vulnerabilities of observations; secondly, it learns the three kinds of couplings across homogeneous and heterogeneous markets, which serve as the key factors driving market dynamics. More specifically, CTDBN outperforms CHMM as CTDBN relies on a deep architecture, which learns inter-market coupling from low-level intra-market coupling, while CHMM cannot. In addition, CTDBN outperforms its sub-model CGRBM, because CGRBM does not built inter-market coupling, it only simply considers the intra-market coupling and temporal coupling.

Figure 7.5 plots the precision and recall of all comparative approaches in the US stock market, where the horizontal axis stands for the number of predicted trading weeks in upward trends, and the vertical axis represents the values of technical measures. We can see that our CTDBN outperforms all other comparative methods. For example, precision improvement in Figure. 7.5 (a) could be as high as 20% against the ARIMA approach, and around 5% against the CHMM and CGRBM methods when k equals to 75. And Figure 7.5 (b) shows the CTDBN achieve higher recall than other models with any number of predicted trading weeks.

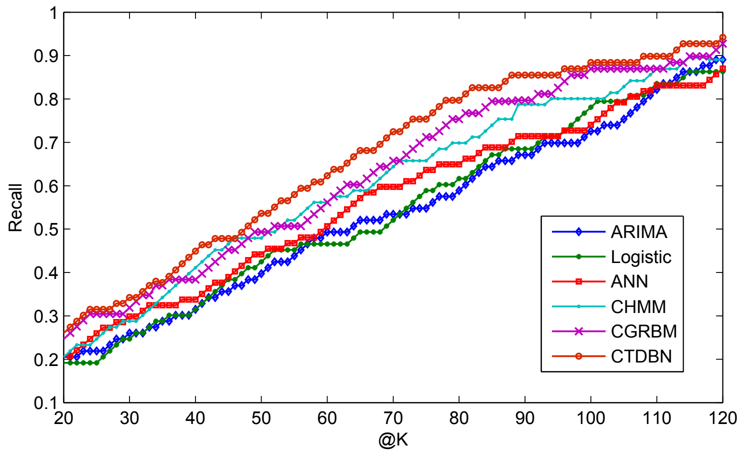
7.5 Summary

In this chapter, we propose a deep learning approach to capture the underlying complex couplings across multiple financial markets. Our model aims to learn hidden features and capture complex couplings across markets. The empirical results of trading the market trends predicted by the model in real financial market show that the proposed approach achieves better outcomes compared to the state-of-the-art methods, from technique and business perspectives.

Obviously, CTDBN has the potential to capture couplings within other inter-dependent scenarios. As CTDBN is a general temporal model, it can be applied to model temporal data, such as human motion (Taylor et al. 2006), music generation (Boulanger-Lewandowski, Bengio & Vincent 2012) and so on. We also further test our model for learning coupled group/community behaviors which are widely seen in the real world but very challenging to model.



(a) Precision



(b) Recall

Figure 7.5: Precision and Recall of Comparative Methods

Chapter 8

Conclusions and Future Work

8.1 Conclusions

Cross-market analysis is not a trivial task since it has non-linear and dynamic features. More importantly, the financial markets interact with each other under particular kinds of coupled behaviors, namely the effect of one market will transfer to another and this is driven by the changes in the hidden coupled behaviors. This has been verified through the financial crisis period. All these have led to the need for and emergence of modeling complex coupled behaviors between different financial markets for typical cross-market problems including financial crisis detection and market trend forecasting.

However, there are various coupled structures of financial markets which have contributed to challenges in understanding and modeling them. For example, the coupled behaviors between different types of markets (e.g. the stock market and commodity) and the same type of market in different countries (e.g. the US and Chinese stock markets). In addition, in order to capture complex coupled behaviors, three major types of couplings need to be explored: intra-coupling, inter-coupling and temporal coupling. Moreover, the couplings are driven by hidden features which cannot be found directly from observation/data. Further, with different cross-market problems, the selection of input factors need to be paid attention to.

Based on this, we present several techniques to understand and model the complex coupled behaviors across financial markets for cross-market analysis.

Chapters 3 and 4 focus on financial crisis detection problems. In Chapter 3, we propose a framework of coupled market behavior analysis, to detect different coupled behaviors between different global financial markets at crisis and non-crisis periods. Here the coupled behaviors refer to the behaviors of different market instruments, such as the stock market index and gold price, which show strong coupled relations. The framework works on the assumption that financial crisis is better reflected through coupled market behaviors rather than single indicators, and the couplings change with the occurrence of the financial crisis, in particular they share different coupling dynamics during and outside the crisis period. We first verify the assumption through a case study from a quantitative perspective, then CHMM is explored to capture and model the coupled behaviors in the two periods. After that, we detect the financial crisis by observing the significant difference which occurs between the crisis and non-crisis periods. The motivation of Chapter 4 is to overcome the limitation of current approaches which forecast the financial crisis directly from observation, while overlooking the hidden driven features. In this chapter we investigate and forecast the financial crisis through teasing out the coupled relationships between hidden market states. The coupled market states here refer to a set of dynamic hidden states from different types of global financial markets, which are used to represent the hidden state transitions generated by interactions between different markets. Accordingly, a Coupled State Space Model is built to capture the coupled market states which are fed into current approaches as features. The performance comparison of our proposed method with current observation-based methods reflects the effect of the coupled market state analysis.

Chapter 5 to Chapter 7 conducts market trend forecasting with different coupled structures and coupling types. In Chapter 5, we predict one global market (e.g. the global commodity market) trends through teasing out its coupled relations with other related global markets such as the US stock market and interest market. A framework of coupled cross-market behavior analysis is proposed to capture the complex couplings among the market behaviors. Specifically, with the help of CHMM, we infer the market trend through forecasting its probability distribution. Chapter 6 pays attention to the hierarchical coupled behaviors between different types of markets in various countries. Here, hierarchical cross-market behaviors refer to the interactions between different types of markets in one country, namely intra-country market coupling (Layer-1 coupling), and the interactions between the same type of markets in different countries which are called inter-country market coupling (Layer-2 coupling). A multi-layer CHMM is built to capture the complex hierarchical coupled relations, in which the Layer-1 coupling is modeled by a CHMM, and the hierarchical interactions between different countries are captured by the multi-layer CHMM. After this we infer a stock market trend by estimating its price return probabilities through the means of capturing the hierarchical coupled market behaviors which are captured within a multi-layer CHMM. Experiments with real financial data in stock and currency markets in 13 major countries verify the advantages of our proposed method. To deep modeling the hidden features which drive the couplings, Chapter 7 proposes a deep architecture, called coupled temporal deep belief network, to capture high-level coupled features. The approach encodes three types of couplings: intra-market coupling represents the interactions between homogeneous markets (e.g. the UK stock market and Chinese stock market), inter-market coupling indicates the interactions between heterogeneous markets (e.g. the US currency market and stock market), and temporal coupling describes the transitional influence across different time points in a market. Formally, we employ the Conditional Gaussian Restricted Boltzmann Machines to learn the abstract features to represent intra-market interactions and the corresponding temporal dependence between homogeneous markets in the first layer. In the second layer, Coupled Conditional Restricted Boltzmann Machines are built on the features learned from the first-layer models, so as to capture the high-level inter-market coupling between heterogeneous markets. Experimental results in relation to the model's capacity to predict market trends in the stock and currency markets of three countries show that the proposed approach achieves better outcomes compared to existing state-of-the-art methods.

Each chapter (i.e. from Chapter 3 to Chapter 7) of this thesis is supported by one published or submitted paper[1] listed in the **List of Publications**. Therefore, what we have done and proposed in this thesis is of real significance to the research into understanding and modeling the coupled behaviors across financial markets and their application to cross-market analysis.

8.2 Future Work

In this section, our ongoing work and the future directions of modeling coupled behaviors across financial markets are listed and stated in terms of three perspectives:

- From the data perspective, more types of data should be taken into consideration. In our thesis, we mainly focus on the closing price as an indicator of each financial market. However, in a real financial market, there exist other kinds of indicators. For example, for financial crisis detection, the political news represents an important indicator in the market since it plays pivotal role in market behaviors, and geographical semantics between countries should be considered in modeling coupled behaviors between various countries. In addition, in relation to market trends forecasting, the coupled behaviors may not only be reflected by the market closing price. Indeed, if we incorporate other indicators such as the opening price and trading volume, we may achieve a better performance.

- From the methodology perspective, one future direction is to build models with more precise predictions. As for market trend forecasting,

[1]The papers of chapter 3, 4 and 7 are published, the papers of chapter 5 and 6 are under review.

we only consider upward and downward directions. It would be more meaningful to give extended directions such as 5% up or 3% down. In addition, it is important to consider whether we can build more expressive but not complex models to capture the various couplings between financial markets. For example, Chapter 6 proposes a multi-layer CHMM to capture couplings between different markets in various countries, however, if many markets and countries are involved in the model, the computation would be too complex. One possible solution is to begin by clustering the similar markets and countries. Also, the complexity and efficiency of the models should be considered, especially when we involve high frequency data (i.e. tick-by-tick data) which is really useful in real financial markets.

- From the application perspective, portfolio optimization, as an important area in cross-market analysis, can be viewed as one future application. This is because it is very important for investors to choose their preferred portfolios, depending on individual risk tolerance. In addition, as illustrated in Chapter 7, coupled behavior analysis can be investigated in learning coupled group/community behaviors which are widely seen in the real world but are very challenging to model.

Chapter 9

List of Symbols

The following list is neither exhaustive nor exclusive, but may be helpful.

9.1 Chapter 3-Chapter 6

PI_t	Price of indicator at time t
RI_t	Return of indicator at time t
I	Number of financial markets
MIC	Market indicator correlation
K	Number of countries
CIC	Country index correlation
$FM(\cdot)$	Behavior Feature Matrix
$\theta(\cdot)$	Intra-coupling function
$\eta(\cdot)$	Inter-coupling function
$f(\theta(\cdot), \eta(\cdot))$	Coupling function
$g(\cdot)$	Objective function

$\Phi(\cdot)$	Market sequence
$CHMM$	Coupled Hidden Markov Model
$MCHMM$	Multi-layer Coupled Hidden Markov Model
$\Omega = (A, B, R, \pi)$	Parameters of CHMM
$\Omega = (A, B, CR, \pi)$	Parameters of MCHMM
A	State transition probability matrix
B	Observation probability matrix
R	Coupling coefficient
π	Initial state probability distribution
$CR = \{CR_1, CR_2\}$	Coupling weight
$\{Z_1, Z_2, \ldots, Z_H\}$	A set of hidden states
z_t	The hidden state at time t
$\{X_1, X_2, \ldots, X_V\}$	A set of observation symbols
$\{O_1, O_2, \ldots, O_T\}$	An observation sequence
o_t	The observation at time t
ANN	Artificial Neural Network
MLE	Maximum Likelihood Estimator
$ARIMA$	Autoregressive Integrated Moving Average
ARR	Annualized Rate of Return
ROR	Rate of Return

9.2 Chapter 7

$\boldsymbol{\theta}_i$	Intra-market coupling w.r.t. the market i
$\boldsymbol{\eta}$	Inter-market coupling
$\boldsymbol{\theta}_{i,t}\lvert\{m_{ij,[t-n,t-1]}\}_{j=1}^{J}$	n-order temporal coupling w.r.t. $\boldsymbol{\theta}_i$
$\boldsymbol{\eta}_t\lvert\{\boldsymbol{\theta}_{i,[t-n,t-1]}\}_{i=1}^{I}$	n-order temporal coupling w.r.t. $\boldsymbol{\eta}$
$\mathbf{v}$	A vector of binary visible units
$\mathbf{h}$	A vector of binary hidden units
$\mathbf{u}$	A vector of binary visible units
$E(\cdot)$	Energy function
$CRBM$	Conditional Restricted Boltzmann Machines
$\Omega = \{\mathbf{W}, \mathbf{A}, \mathbf{B}, \mathbf{a}, \mathbf{b}\}$	Parameters of CRBM
$\mathbf{W}$	Visible-hidden interactions between $\mathbf{v}$ and $\mathbf{h}$
$\mathbf{A}$	Interactions between $\mathbf{u}$ and $\mathbf{v}$
$\mathbf{B}$	Interactions between $\mathbf{u}$ and $\mathbf{h}$
$\mathbf{a}$	Bias of $\mathbf{v}$
$\mathbf{b}$	Bias of $\mathbf{h}$
$\mathbf{m}_t$	Market observations at time t
$\mathbf{m}_{<t}$	A stacked history vector of market observations

Bibliography

Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. (1985), 'A learning algorithm for boltzmann machines*', *Cognitive science* **9**(1), 147–169.

Ahmed, A., Yu, K., Xu, W., Gong, Y. & Xing, E. (2008), Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks, *in* 'Computer Vision–ECCV 2008', Springer, pp. 69–82.

Aktan, B., Kopurlu, B. S. et al. (2009), 'Behaviour of emerging stock markets in the global financial meltdown: Evidence from bric-a', *African Journal of Business Management* **3**(9), 396–404.

Alvarez-Plata, P. & Schrooten, M. (2004), 'Misleading indicators? the argentinean currency crisis', *Journal of Policy Modeling* **26**(5), 587–603.

Atiya, A. F. (2001), 'Bankruptcy prediction for credit risk using neural networks: A survey and new results', *Neural Networks, IEEE Transactions on* **12**(4), 929–935.

Ayres, J., Flannick, J., Gehrke, J. & Yiu, T. (2002), Sequential pattern mining using a bitmap representation, *in* 'Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 429–435.

Baesens, B., Setiono, R., Mues, C. & Vanthienen, J. (2003), 'Using neural network rule extraction and decision tables for credit-risk evaluation', *Management Science* **49**(3), 312–329.

Baker, M. J. & Hart, S. J. (2008), *The marketing book*, Routledge.

Baldi, P. & Brunak, S. (2001), *Bioinformatics: the machine learning approach*, MIT press.

Batal, I., Sacchi, L., Bellazzi, R. & Hauskrecht, M. (2009), Multivariate time series classification with temporal abstractions., *in* 'FLAIRS Conference'.

Baum, L. E., Eagon, J. et al. (1967), 'An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology', *Bull. Amer. Math. Soc* **73**(3), 360–363.

Baum, L. E. & Petrie, T. (1966), 'Statistical inference for probabilistic functions of finite state markov chains', *The annals of mathematical statistics* pp. 1554–1563.

Baum, L. E., Petrie, T., Soules, G. & Weiss, N. (1970), 'A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains', *The annals of mathematical statistics* pp. 164–171.

Baum, L. E. & Sell, G. (1968), 'Growth functions for transformations on manifolds', *.¡. Math* **27**(2), 211–227.

Beckmann, D., Menkhoff, L. & Sawischlewski, K. (2006), 'Robust lessons about practical early warning systems', *Journal of Policy Modeling* **28**(2), 163–193.

Bengio, Y. (2009), 'Learning deep architectures for ai', *Foundations and trends® in Machine Learning* **2**(1), 1–127.

Bengio, Y., Courville, A. & Vincent, P. (2013), 'Representation learning: A review and new perspectives', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **35**(8), 1798–1828.

Bengio, Y. & Frasconi, P. (1996), 'Input-output hmms for sequence processing', *Neural Networks, IEEE Transactions on* **7**(5), 1231–1249.

Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. (2007), 'Greedy layer-wise training of deep networks', *Advances in neural information processing systems* **19**, 153.

Bengio, Y. & LeCun, Y. (2007), 'Scaling learning algorithms towards ai', *Large-scale kernel machines* **34**, 1–41.

Berg, A. & Pattillo, C. (1999), 'Predicting currency crises:: The indicators approach and an alternative', *Journal of International Money and Finance* **18**(4), 561–586.

Bickhard, M. (2000), 'An interactive process model', *The caldron of consciousness: motivation, affect, and self-organization: an an thology* p. 161.

Bodart, V. & Reding, P. (1999), 'Exchange rate regime, volatility and international correlations on bond and stock markets', *Journal of International Money and Finance* **18**(1), 133–151.

Boreczky, J. S. & Wilcox, L. D. (1998), A hidden markov model framework for video segmentation using audio and image features, *in* 'Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on', Vol. 6, IEEE, pp. 3741–3744.

Boulanger-Lewandowski, N., Bengio, Y. & Vincent, P. (2012), 'Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription', *arXiv preprint arXiv:1206.6392* .

Box, G. E., Jenkins, G. M. & Reinsel, G. C. (2013), *Time series analysis: forecasting and control*, John Wiley & Sons.

Box, G. E. & Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American statistical Association* **65**(332), 1509–1526.

Brand, M. (1997), 'Coupled hidden markov models for modeling interacting processes'.

Brand, M., Oliver, N. & Pentland, A. (1997), Coupled hidden markov models for complex action recognition, *in* 'Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on', IEEE, pp. 994–999.

Brunnermeier, M. K. (2008), Deciphering the liquidity and credit crunch 2007-08, Technical report, National Bureau of Economic Research.

Bussière, M. & Fratzscher, M. (2006), 'Towards a new early warning system of financial crises', *journal of International Money and Finance* **25**(6), 953–973.

Candelon, B., Piplack, J. & Straetmans, S. (2009), 'Multivariate business cycle synchronization in small samples*', *Oxford Bulletin of Economics and Statistics* **71**(5), 715–737.

Cao, L. (2010), 'In-depth behavior understanding and use: the behavior informatics approach', *Information Sciences* **180**(17), 3067–3085.

Cao, L., Ou, Y. & Yu, P. S. (2012), 'Coupled behavior analysis with applications', *Knowledge and Data Engineering, IEEE Transactions on* **24**(8), 1378–1392.

Cao, L., Ou, Y., Yu, P. S. & Wei, G. (2010), Detecting abnormal coupled sequences and sequence changes in group-based manipulative trading behaviors, *in* 'Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 85–94.

Cao, L. & Tay, F. E. (2001), 'Financial forecasting using support vector machines', *Neural Computing & Applications* **10**(2), 184–192.

Celik, A. E. & Karatepe, Y. (2007), 'Evaluating and forecasting banking crises through neural network models: An application for turkish banking sector', *Expert systems with Applications* **33**(4), 809–815.

Chan, K. F., Treepongkaruna, S., Brooks, R. & Gray, S. (2011), 'Asset market linkages: Evidence from financial, commodity and real estate assets', *Journal of Banking & Finance* **35**(6), 1415–1426.

Chandrakala, S. & Sekhar, C. C. (2008), A density based method for multivariate time series clustering in kernel feature space, *in* 'Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on', IEEE, pp. 1885–1890.

Chatfield, C. (2013), *The analysis of time series: an introduction*, CRC press.

Chen, C., Liang, J., Zhao, H., Hu, H. & Tian, J. (2009), 'Factorial hmm and parallel hmm for gait recognition', *IEEE Transactions on Systems, Man, and Cybernetics, Part C* **39**(1), 114–123.

Chen, G.-m., Firth, M. & Meng Rui, O. (2002), 'Stock market linkages: evidence from latin america', *Journal of Banking & Finance* **26**(6), 1113–1141.

Chen, S. M. & Chen, C. D. (2011), 'Taiex forecasting based on fuzzy time series and fuzzy variation groups', *Fuzzy Systems, IEEE Trans on* **19**(1), 1–12.

Chevallier, J. (2012), 'Global imbalances, cross-market linkages, and the financial crisis: A multivariate markov-switching analysis', *Economic Modelling* **29**(3), 943–973.

Contreras, J., Espinola, R., Nogales, F. J. & Conejo, A. J. (2003), 'Arima models to predict next-day electricity prices', *Power Systems, IEEE Trans on* **18**(3), 1014–1020.

Dai, R., Zhang, C. & Cao, L. (2005), 'Organization-oriented analysis of open complex agent systems'.

Daskalaki, C. & Skiadopoulos, G. (2011), 'Should investors include commodities in their portfolios after all? new evidence', *Journal of Banking & Finance* **35**(10), 2606–2626.

de Souza e Silva, E. G., Legey, L. F. & de Souza e Silva, E. A. (2010), 'Forecasting oil price trends using wavelets and hidden markov models', *Energy Economics* **32**(6), 1507 – 1519.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–38.

Durbin, J. & Koopman, S. J. (2012), *Time series analysis by state space methods*, number 38, Oxford University Press.

Eichengreen, B., Rose, A. K., Wyplosz, C., Dumas, B. & Weber, A. (1995), 'Exchange market mayhem: the antecedents and aftermath of speculative attacks', *Economic policy* pp. 249–312.

Elliott, R. J., Aggoun, L. & Moore, J. B. (1994), *Hidden Markov Models*, Springer.

Fast, A., Friedland, L., Maier, M., Taylor, B., Jensen, D., Goldberg, H. G. & Komoroske, J. (2007), Relational data pre-processing techniques for improved securities fraud detection, *in* 'Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 941–949.

Fine, S., Singer, Y. & Tishby, N. (1998), 'The hierarchical hidden markov model: Analysis and applications', *Machine learning* **32**(1), 41–62.

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes*, Springer.

Ghahramani, Z. & Hinton, G. E. (1996), Parameter estimation for linear dynamical systems, Technical report, Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science.

Ghahramani, Z. & Hinton, G. E. (2000), 'Variational learning for switching state-space models', *Neural computation* **12**(4), 831–864.

Ghahramani, Z. & Jordan, M. I. (1997), 'Factorial hidden markov models', *Machine learning* **29**(2-3), 245–273.

Godsill, S., Doucet, A. & West, M. (2000), 'Methodology for monte carlo smoothing with application to time-varying autoregressions'.

Goebel, R., Roebroeck, A., Kim, D.-S. & Formisano, E. (2003), 'Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping', *Magnetic resonance imaging* **21**(10), 1251–1261.

Goldstein, M., Kaminsky, G. L. & Reinhart, C. M. (2000), *Assessing financial vulnerability: an early warning system for emerging markets*, Peterson Institute.

Goodwin, G. C. & Sin, K. S. (2013), *Adaptive filtering prediction and control*, Courier Dover Publications.

Hadsell, R., Erkan, A., Sermanet, P., Scoffier, M., Muller, U. & LeCun, Y. (2008), Deep belief net learning in a long-range vision system for autonomous off-road driving, *in* 'Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on', IEEE, pp. 628–633.

Halliday, R. (2004), 'Equity trend prediction with neural networks'.

Hamori, S. & Imamura, Y. (2000), 'International transmission of stock prices among g7 countries: La-var approach', *Applied Economics Letters* **7**(9), 613–618.

Harshman, R. A. & Lundy, M. E. (1994), 'Parafac: Parallel factor analysis', *Computational Statistics & Data Analysis* **18**(1), 39–72.

Hassan, M. R. & Nath, B. (2005), Stock market forecasting using hidden markov model: a new approach, *in* 'Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference on', IEEE, pp. 192–196.

Hassan, M. R., Nath, B. & Kirley, M. (2007), 'A fusion model of hmm, ann and ga for stock market forecasting', *Expert Systems with Applications* **33**(1), 171–180.

Haykin, S. S., Haykin, S. S. & Haykin, S. S. (2001), *Kalman filtering and neural networks*, Wiley Online Library.

Hinton, G. E. & Salakhutdinov, R. (2008), Using deep belief nets to learn covariance kernels for gaussian processes, *in* 'Advances in neural information processing systems', pp. 1249–1256.

Hinton, G. E. & Sejnowski, T. J. (1986), 'Learning and relearning in boltzmann machines', *Cambridge, MA: MIT Press* **1**(282-317), 4–2.

Hinton, G., Osindero, S. & Teh, Y.-W. (2006), 'A fast learning algorithm for deep belief nets', *Neural computation* **18**(7), 1527–1554.

Hodge, V. J. & Austin, J. (2004), 'A survey of outlier detection methodologies', *Artificial Intelligence Review* **22**(2), 85–126.

Holcombe, R. G. (1989), 'Economic models and methodology'.

Hosmer Jr, D. W. & Lemeshow, S. (2004), *Applied logistic regression*, John Wiley & Sons.

Huarng, K. & Yu, T. H.-K. (2006), 'The application of neural networks to forecast fuzzy time series', *Physica A: Statistical Mechanics and its Applications* **363**(2), 481–491.

Hurvich, C. M. & Tsai, C.-L. (1989), 'Regression and time series model selection in small samples', *Biometrika* **76**(2), 297–307.

Hyup Roh, T. (2007), 'Forecasting the volatility of stock price index', *Expert Systems with Applications* **33**(4), 916–922.

Ilgen, D. R., Hulin, C. L., Association, A. P. et al. (2000), *Computational modeling of behavior in organizations: The third scientific discipline*, American Psychological Association Washington.

Jang, W., Lee, J. & Chang, W. (2011), 'Currency crises and the evolution of foreign exchange market: Evidence from minimum spanning tree', *Physica A: Statistical Mechanics and its Applications* **390**(4), 707–718.

Johansen, S. & Juselius, K. (1990), 'Maximum likelihood estimation and inference on cointegrationwith applications to the demand for money', *Oxford Bulletin of Economics and statistics* **52**(2), 169–210.

Jordan, M. I. (2002), 'An introduction to probabilistic graphical models'.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Fluids Engineering* **82**(1), 35–45.

Kalman, R. E. & Bucy, R. S. (1961), 'New results in linear filtering and prediction theory', *Journal of Fluids Engineering* **83**(1), 95–108.

Kalotychou, E. & Staikouras, S. K. (2006), 'An empirical investigation of the loan concentration risk in latin america', *Journal of Multinational Financial Management* **16**(4), 363–384.

Kaminsky, G. L. (1999), *Currency and banking crises-the early warnings of distress*, International Monetary Fund.

Kaminsky, G. L. & Reinhart, C. M. (1999), 'The twin crises: the causes of banking and balance-of-payments problems', *American economic review* pp. 473–500.

Kim, C.-J. (1994), 'Dynamic linear models with markov-switching', *Journal of Econometrics* **60**(1), 1–22.

Kim, T. Y., Hwang, C. & Lee, J. (2004), 'Korean economic condition indicator using a neural network trained on the 1997 crisis', *Journal of Data Science* **2**(4), 371–381.

Kim, T. Y., Oh, K. J., Sohn, I. & Hwang, C. (2004), 'Usefulness of artificial neural networks for early warning system of economic crisis', *Expert Systems with Applications* **26**(4), 583–590.

Kitagawa, G. (1998), 'A self-organizing state-space model', *Journal of the American Statistical Association* pp. 1203–1215.

Koller, D. & Friedman, N. (2009), *Probabilistic graphical models: principles and techniques*, MIT press.

Kumar, M., Moorthy, U. & Perraudin, W. (2003), 'Predicting emerging market currency crashes', *Journal of Empirical Finance* **10**(4), 427–454.

Laitinen, E. K. & Laitinen, T. (2001), 'Bankruptcy prediction: application of the taylor's expansion in logistic regression', *International Review of Financial Analysis* **9**(4), 327–349.

Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. (2009), Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, *in* 'Proceedings of the 26th Annual International Conference on Machine Learning', ACM, pp. 609–616.

Levinson, S. E., Rabiner, L. R. & Sondhi, M. M. (1983), 'An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition', *Bell System Technical Journal* **62**(4), 1035–1074.

Lin, C.-S., Khan, H. A., Chang, R.-Y. & Wang, Y.-C. (2008), 'A new approach to modeling early warning systems for currency crises: Can a machine-learning fuzzy expert system predict the currency crises effectively?', *Journal of International Money and Finance* **27**(7), 1098–1121.

Liu, X., Swift, S., Tucker, A., Cheng, G. & Loizou, C. (1999), 'Modelling multivariate time series', *Intelligent Data Analysis in Medicine and Pharmacology* pp. 67–78.

Longstaff, F. A. (2010), 'The subprime credit crisis and contagion in financial markets', *Journal of Financial Economics* **97**(3), 436 – 450.

Ma, F., Wei, Y. & Huang, D. (2013), 'Multifractal detrended cross-correlation analysis between the chinese stock market and surrounding stock markets', *Physica A: Statistical Mechanics and its Applications* **392**(7), 1659–1670.

Marcucci, J. (2005), 'Forecasting stock market volatility with regime-switching garch models', *Studies in Nonlinear Dynamics & Econometrics* **9**(4).

Markov, A. A. (2006), 'An example of statistical investigation of the text eugene onegin concerning the connection of samples in chains', *Science in Context* **19**(04), 591–600.

Mnih, A. & Hinton, G. E. (2009), A scalable hierarchical distributed language model, *in* 'Advances in neural information processing systems', pp. 1081–1088.

Mo, D. (2012), 'A survey on deep learning: one small step toward ai'.

Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E. & Kaushansky, H. (2000), 'Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry', *Neural Networks, IEEE Transactions on* **11**(3), 690–696.

Murphy, K. P. (2002), Dynamic bayesian networks: representation, inference and learning, PhD thesis, University of California, Berkeley.

Natarajan, P. & Nevatia, R. (2007), Coupled hidden semi markov models for activity recognition, *in* 'Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on', IEEE, pp. 10–10.

Nguyen, N. T., Phung, D. Q., Venkatesh, S. & Bui, H. (2005), Learning and detecting activities from movement trajectories using the hierarchical hidden markov model, *in* 'Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on', Vol. 2, IEEE, pp. 955–960.

Oates, T., Schmill, M. D., Cohen, P. R. & Durfee, C. (1999), E cient mining of statistical dependencies, *in* 'Proc. 16th Int. Joint Conf. on Artificial Intelligence', Citeseer, pp. 794–799.

Oliver, N. M., Rosario, B. & Pentland, A. P. (2000), 'A bayesian computer vision system for modeling human interactions', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22**(8), 831–843.

Olson, D. & Mossman, C. (2003), 'Neural network forecasts of canadian stock returns using accounting ratios', *International Journal of Forecasting* **19**(3), 453–465.

Osindero, S. & Hinton, G. E. (2008), Modeling image patches with a directed hierarchy of markov random fields, *in* 'Advances in neural information processing systems', pp. 1121–1128.

OSullivan, A. & Sheffrin, S. M. (2007), *Economics: Principles in action*, Boston, Mass.: Pearson/Prentice Hall.

Ozkan-Gunay, E. N. & Ozkan, M. (2007), 'Prediction of bank failures in emerging financial markets: an ann approach', *Journal of Risk Finance, The* **8**(5), 465–480.

Pan, H., Tilakaratne, C. & Yearwood, J. (2005), 'Predicting australian stock market index using neural networks exploiting dynamical swings and intermarket influences', *Journal of research and practice in information technology* **37**(1), 43–56.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008), 'State–space models of individual animal movement', *Trends in ecology & evolution* **23**(2), 87–94.

Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U. & Hsu, M.-C. (2001), Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth, *in* '2013 IEEE 29th International Conference on Data Engineering (ICDE)', IEEE Computer Society, pp. 0215–0215.

Peng, D. & Bajona, C. (2008), 'China's vulnerability to currency crisis: A klr signals approach', *China Economic Review* **19**(2), 138–151.

Phillips, P. C. & Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**(2), 335–346.

Pierce, W. D. & Cheney, C. D. (2013), *Behavior analysis and learning*, Psychology Press.

Podobnik, B. & Stanley, H. E. (2007), 'Detrended cross-correlation analysis: a new method for analyzing two non-stationary time series', *arXiv preprint arXiv:0709.0281* .

Pole, A., West, M. & Harrison, J. (1994), *Applied Bayesian forecasting and time series analysis*, CRC Press.

Quagrainie, K. K. (2004), Forecasting market share using a flexible logistic model, *in* '2004 Annual Meeting, February 14-18, 2004, Tulsa, Oklahoma', number 34724, Southern Agricultural Economics Association.

Rabiner, L. (1989), 'A tutorial on hidden markov models and selected applications in speech recognition', *Proceedings of the IEEE* **77**(2), 257–286.

Rabiner, L. & Juang, B.-H. (1986), 'An introduction to hidden markov models', *ASSP Magazine, IEEE* **3**(1), 4–16.

Rasmussen, C. E. (1999), The infinite gaussian mixture model., *in* 'NIPS', Vol. 12, pp. 554–560.

Rauch, H. (1963), 'Solutions to the linear smoothing problem', *Automatic Control, IEEE Transactions on* **8**(4), 371–372.

Reinsel, G. C. (2003), *Elements of multivariate time series analysis*, Springer.

Rezek, I., Gibbs, M. & Roberts, S. J. (2002), 'Maximum a posteriori estimation of coupled hidden markov models', *Journal of VLSI signal processing systems for signal, image and video technology* **32**(1-2), 55–66.

Rezek, I. & Roberts, S. J. (2000), Estimation of coupled hidden markov models with application to biosignal interaction modelling, *in* 'Neural Networks Signal Process Proc IEEE', Vol. 2, pp. 804–813.

Sainath, T. N., Kingsbury, B., Ramabhadran, B., Fousek, P., Novak, P. & Mohamed, A.-r. (2011), Making deep belief networks effective for large vocabulary continuous speech recognition, *in* 'Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on', IEEE, pp. 30–35.

Saul, L. K. & Jordan, M. I. (1995), 'Boltzmann chains and hidden markov models', *Advances in neural information processing systems* pp. 435–442.

Sherlock, C., Xifara, T., Telfer, S. & Begon, M. (2013), 'A coupled hidden markov model for disease interactions', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **62**(4), 609–627.

Shin, K.-S., Lee, T. S. & Kim, H.-j. (2005), 'An application of support vector machines in bankruptcy prediction model', *Expert Systems with Applications* **28**(1), 127–135.

Shinghal, R. & Toussaint, G. T. (1979), 'Experiments in text recognition with the modified viterbi algorithm', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (2), 184–193.

Simutis, R. (2000), Fuzzy logic based stock trading system, *in* 'Computational Intelligence for Financial Engineering, 2000.(CIFEr) Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on', IEEE, pp. 19–21.

Söding, J. (2005), 'Protein homology detection by hmm–hmm comparison', *Bioinformatics* **21**(7), 951–960.

Sonnhammer, E. L., Von Heijne, G., Krogh, A. et al. (1998), A hidden markov model for predicting transmembrane helices in protein sequences., *in* 'Ismb', Vol. 6, pp. 175–182.

Srikant, R. & Agrawal, R. (1996), *Mining sequential patterns: Generalizations and performance improvements*, Springer.

Stolyar, A. L. (2004), 'Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic', *Annals of Applied Probability* pp. 1–53.

Sunderlin, W. D., Angelsen, A., Resosudarmo, D. P., Dermawan, A. & Rianto, E. (2001), 'Economic crisis, small farmer well-being, and forest cover change in indonesia', *World Development* **29**(5), 767–782.

Tatavarty, G., Bhatnagar, R. & Young, B. (2007), Discovery of temporal dependencies between frequent patterns in multivariate time series, *in* 'Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on', IEEE, pp. 688–696.

Taylor, G. W. (2009), Composable, distributed-state models for high-dimensional time series, PhD thesis, University of Toronto.

Taylor, G. W. & Hinton, G. E. (2009), Factored conditional restricted boltzmann machines for modeling motion style, *in* 'Proceedings of the 26th annual international conference on machine learning', ACM, pp. 1025–1032.

Taylor, G. W., Hinton, G. E. & Roweis, S. T. (2006), Modeling human motion using binary latent variables, *in* 'Advances in neural information processing systems', pp. 1345–1352.

Teh, Y. W. (2010), Dirichlet process, *in* 'Encyclopedia of machine learning', Springer, pp. 280–287.

Töreyin, B. U., Dedeoğlu, Y. & Çetin, A. E. (2005), Hmm based falling person detection using both audio and video, *in* 'Computer Vision in Human-Computer Interaction', Springer, pp. 211–220.

Tseng, F.-M., Tzeng, G.-H., Yu, H.-C. & Yuan, B. J. (2001), 'Fuzzy arima model for forecasting the foreign exchange market', *Fuzzy sets and systems* **118**(1), 9–19.

Upadhyay, A., Bandyopadhyay, G. & Dutta, A. (2012), 'Forecasting stock performance in indian market using multinomial logistic regression', *Journal of Business Studies Quarterly* **3**(3), 16–39.

Van Hemel, S. B., MacMillan, J., Zacharias, G. L. et al. (2008), *Behavioral Modeling and Simulation:: From Individuals to Societies*, National Academies Press.

Varga, A. & Moore, R. (1990), Hidden markov model decomposition of speech and noise, *in* 'Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on', IEEE, pp. 845–848.

Viterbi, A. J. (1967), 'Error bounds for convolutional codes and an asymptotically optimum decoding algorithm', *Information Theory, IEEE Transactions on* **13**(2), 260–269.

Weigend, A. S. & Shi, S. (1998), 'Predicting daily probability distributions of s&p500 returns'.

Welling, M. & Hinton, G. E. (2002), A new learning algorithm for mean field boltzmann machines, *in* 'Artificial Neural NetworksICANN 2002', Springer, pp. 351–357.

White, H. (1988), Economic prediction using neural networks: The case of ibm daily stock returns, *in* 'Neural Networks, 1988., IEEE International Conference on', IEEE, pp. 451–458.

Yang, J., Min, I. & Li, Q. (2003), 'European stock market integration: does emu matter?', *Journal of business finance & accounting* **30**(9-10), 1253–1276.

Yoon, H., Yang, K. & Shahabi, C. (2005), 'Feature subset selection and feature ranking for multivariate time series', *Knowledge and Data Engineering, IEEE Transactions on* **17**(9), 1186–1198.

Yu, L., Wang, S., Lai, K. K. & Wen, F. (2010), 'A multiscale neural network learning paradigm for financial crisis forecasting', *Neurocomputing* **73**(4), 716–725.

Zhong, S. & Ghosh, J. (2001), A new formulation of coupled hidden markov models, Technical report, Tech. Report, Dept. of Electronic and Computer Engineering, U. of Texas at Austin, USA.

Zhong, S. & Ghosh, J. (2002), Hmms and coupled hmms for multi-channel eeg classification, *in* 'Proceedings of the IEEE International Joint Conference on Neural Networks', Vol. 2, pp. 1254–1159.