Deep Modeling Complex Couplings within Financial Markets

Wei Cao
Advanced Analytics Institute
University of Technology, Sydney
Wei.Cao@student.uts.edu.au

Liang Hu
University of Technology, Sydney
and Shanghai Jiaotong University
lianghu@sjtu.edu.cn

Longbing Cao
Advanced Analytics Institute
University of Technology, Sydney
LongBing.Cao@uts.edu.au

Abstract

The global financial crisis occurred in 2008 and its contagion to other regions, as well as the long-lasting impact on different markets, show that it is increasingly important to understand the complicated coupling relationships across financial markets. This is indeed very difficult as complex hidden coupling relationships exist between different financial markets in various countries, which are very hard to model. The couplings involve interactions between homogeneous markets from various countries (we call *intra-market coupling*), interactions between heterogeneous markets (*inter-market coupling*) and interactions between current and past market behaviors (*temporal coupling*). Very limited work has been done towards modeling such complex couplings, whereas some existing methods predict market movement by simply aggregating indicators from various markets but ignoring the inbuilt couplings. As a result, these methods are highly sensitive to observations, and may often fail when financial indicators change slightly. In this paper, a coupled deep belief network is designed to accommodate the above three types of couplings across financial markets. With a deep-architecture model to capture the high-level coupled features, the proposed approach can infer market trends. 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.

Introduction

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 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 *temporary 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 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.

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 and 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 and 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

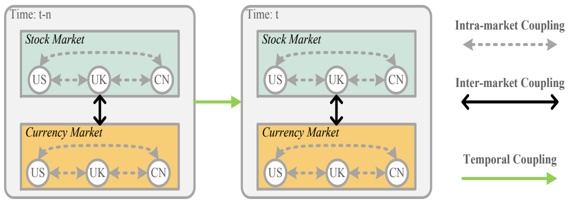


Figure 1: A demonstration of complex couplings between financial markets

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 paper, 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 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.

Related Work

Here we mainly discuss the approaches related to time series analysis and machine learning-based models, which are widely used for cross-market analysis.

Time Series Models

In financial markets, time series analysis uses historical data to infer future trend behaviors; typical representatives are ARIMA and Logistic models. The linear ARIMA model was used in (Contreras et al. 2003) to analyze time series from mainland Spain and California markets to predict next-day electricity market price. The Logistic model was used in (Laitinen and Laitinen 2001) to solve the bankruptcy prediction problem. In (Chen and Chen 2011), fuzzy multivariate time series analysis was used to forecast the daily Taiwan stock index. However, this kind of method suffer from market changes of the corresponding input 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.

Machine Learning Models

Machine learning-based models have been increasingly explored for financial market analysis. The typical models include ANN (Olson and Mossman 2003; Pan, Tilakaratne, and Yearwood 2005) and HMM (Hassan and Nath 2005; Cao et al. 2010), which check any systematic patterns in time series for prediction. In (Huarng and Yu 2006), ANN was used to establish fuzzy relationships in fuzzy time series for stock forecasting, and the volatility of stock price index is predicted by ANN in (Hyup Roh 2007). In (Hassan and Nath 2005), HMM was used to forecast stock price for interrelated markets. Although these methods explored temporal couplings in a market or correlations between markets, they do not effectively address the intra- and inter-market couplings between homogeneous and heterogeneous markets with multiple distinct financial indicators. Recently, Coupled Hidden Markov Model (CHMM) was deployed to learn the coupled market behaviors (Cao, Ou, and Yu 2012), which captures the hidden couplings between multiple time series. However, the modeling structure of CHMM is not powerful enough to model the complex cross-market couplings due to its structural limitation.

Preliminaries

We introduce some concepts used in this paper 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.

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 paper, we focus on representing three types of couplings (cf. Figure 1): *intra-market coupling*, *inter-market coupling* and *temporal coupling*. The corresponding definitions are as follows:

Definition 1. Intra-market Coupling : This is the interaction between homogeneous markets from all countries. Formally, the representation of intra-market interaction w.r.t the market i is given by:

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

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

Definition 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:

$$\boldsymbol{\eta} = \{\circledast_{i=1}^{I}(\boldsymbol{\theta}_i)\} \tag{2}$$

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

Definition 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:

$$\boldsymbol{\theta}_{i,t}|\{m_{ij,[t-n,t-1]}\}_{j=1}^{J} \tag{3}$$

$$\boldsymbol{\eta}_t|\{\boldsymbol{\theta}_{i,[t-n,t-1]}\}_{i=1}^{I} \tag{4}$$

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

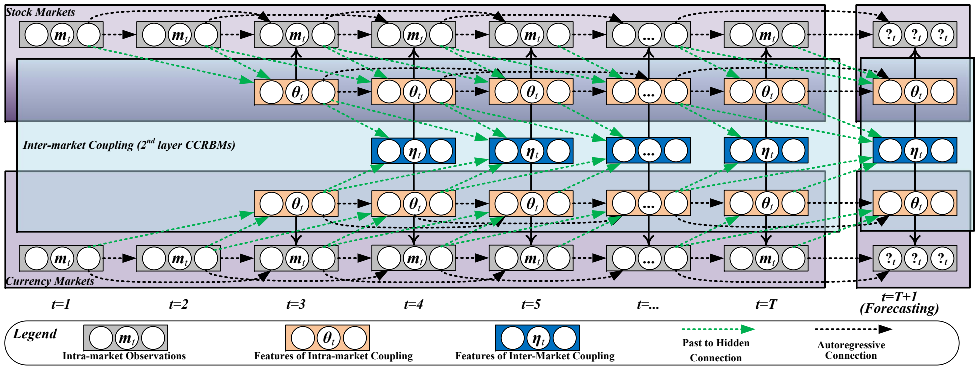


Figure 2: Modeling framework of CTDBN. Here, the demonstration shows two heterogeneous financial markets, stock and currency. The first-layer are CGRBMs to model the intra-maket couplings while CCRBMs are built on the first layer to model inter-market couplings.

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

$$P(\mathbf{v},\mathbf{h} \mid \mathbf{u}) = exp(-E(\mathbf{v},\mathbf{h},\mathbf{u}))/Z \tag{5}$$

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

$$E(\mathbf{v},\mathbf{h},\mathbf{u}) = -\mathbf{v}^{\mathbf{T}}\mathbf{W}\mathbf{h} - \mathbf{u}^{\mathbf{T}}\mathbf{A}\mathbf{v} - \mathbf{u}^{\mathbf{T}}\mathbf{B}\mathbf{h} - \mathbf{a}^{\mathbf{T}}\mathbf{v} - \mathbf{b}^{\mathbf{T}}\mathbf{h} \tag{6}$$

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, Courville, and Vincent 2013), which can easily derived from Eq. (5):

$$P(h_f = 1 \mid \mathbf{u},\mathbf{v}) = s(b_f + \mathbf{u}^{\mathbf{T}}\mathbf{B}_{:,f} + \mathbf{v}^{\mathbf{T}}\mathbf{W}_{:,f}) \tag{7}$$

$$P(v_d = 1 \mid \mathbf{v},\mathbf{u}) = s(a_d + \mathbf{u}^{\mathbf{T}}\mathbf{A}_{:,d} + \mathbf{W}_{d,:}\mathbf{h}) \tag{8}$$

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 paper.

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:

$$E(\mathbf{v},\mathbf{h},\mathbf{u}) = \tag{9}$$

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

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

$$P(h_f = 1 \mid \mathbf{v},\mathbf{u}) = s(b_f + \mathbf{u}^{\mathbf{T}}\mathbf{B}_{:,f} + \mathbf{v}^{\mathbf{T}}\mathbf{W}_{:,f}/\sigma) \tag{10}$$

$$P(v_d \mid \mathbf{v},\mathbf{u}) = \mathcal{N}(a_d + \mathbf{u}^{\mathbf{T}}\mathbf{A}_{:,d} + \sigma\mathbf{W}_{d,:}\mathbf{h},\sigma^2) \tag{11}$$

In many applications, it is much easier to normalize each visible unit to zero mean and unit variance (Taylor and Hinton 2009), so that we can simply set $\sigma = 1$. In this paper, 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 $\omega \in \Omega$:

$$-\frac{\partial logp(\mathbf{v})}{\partial\omega} = \mathbb{E}_{P(\mathbf{h}|\mathbf{v},\mathbf{u})}(\frac{\partial E(\mathbf{v},\mathbf{h},\mathbf{u})}{\partial\omega}) - \mathbb{E}_{P(\mathbf{v},\mathbf{h},\mathbf{u})}(\frac{\partial E(\mathbf{v},\mathbf{h},\mathbf{u})}{\partial\omega}) \tag{12}$$

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, Courville, and Vincent 2013). In practice, Contrastive Divergence (CD) (Hinton, Osindero, and Teh 2006) can be used to approximate the expectation with a short $k - step$ (e.g. $k = 1$) Gibbs sampling using Eq. (7), (8) and (10), (11), denoted as CD_k.

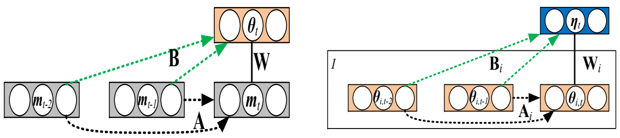
The stochastic gradient descent update using CD_k w.r.t. each parameter $\omega \in \Omega$ can be given by:

$$\omega \leftarrow \omega - \alpha(\frac{\partial E(\mathbf{v}^0,\mathbf{h}^0,\mathbf{u})}{\partial\omega} - \frac{\partial E(\mathbf{v}^k,\mathbf{h}^k,\mathbf{u})}{\partial\omega}) \tag{13}$$

where $\mathbf{v}^0$ are the visible data, $\mathbf{h}^0$ is sampled by Eq. (7) or (10), $\mathbf{v}^k$ and $\mathbf{h}^k$ are sampled from the k-step Gibbs chain.

Modeling and Forecasting

In this section, we focus on modeling the coupled relations in the global financial markets as formalized by Definitions 1, 2 and 3. We design a CTDBN to hierarchically model such complex interactions. Figure 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.



(a) (b)

Figure 3: (a) A CGRBM to model intra-market coupling at time t; (b) A CCRBM to model inter-market coupling at time t

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

As shown in Figure 3 (a), 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 Eq. (9).

$$E(\mathbf{m}_t, \boldsymbol{\theta}_t, \mathbf{m}_{<t}) = \frac{(\mathbf{m}_t - \mathbf{a})^T(\mathbf{m}_t - \mathbf{a})}{2} - \mathbf{m}_t^T\boldsymbol{\theta}_t - \mathbf{m}_{<t}^T\mathbf{A}\mathbf{m}_t - \mathbf{m}_{<t}^T\mathbf{B}\boldsymbol{\theta}_t - \mathbf{b}^T\boldsymbol{\theta}_t \tag{14}$$

Then the conditional distributions can be immediately obtained according to Eq. (10) and (11).

$$P(\theta_{ft} = 1 \mid \mathbf{m}_t, \mathbf{m}_{<t}) = s(b_f + \mathbf{m}_{<t}^T\mathbf{B}_{:,f} + \mathbf{m}_t^T\mathbf{W}_{:,f}) \tag{15}$$

$$P(m_{jt} \mid \boldsymbol{\theta}_t, \mathbf{m}_{<t}) = \mathcal{N}(a_j + +\mathbf{m}_{<t}^T\mathbf{A}_{:,j} + \mathbf{W}_{i,:}\boldsymbol{\theta}_t, 1) \tag{16}$$

Formalism Mapping We easily find that the *intra-market coupling* operator $\otimes$ in Eq. (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*.

Parameter Learning Given Eq. (14), (15) and (16), the stochastic gradient update equations using CD_k can be obtained according to Eq. (13).

$$\mathbf{W} \leftarrow \mathbf{W} - \alpha(\mathbf{m}_t^{(0)}\boldsymbol{\theta}_t^{(0)T} - \mathbf{m}_t^{(k)}\boldsymbol{\theta}_t^{(k)T}) \tag{17}$$

$$\mathbf{A} \leftarrow \mathbf{A} - \alpha(\mathbf{m}_{<t}^{(0)}\mathbf{m}_t^{(0)T} - \mathbf{m}_{<t}^{(k)}\mathbf{m}_t^{(k)T}) \tag{18}$$

$$\mathbf{B} \leftarrow \mathbf{B} - \alpha(\mathbf{m}_{<t}^{(0)}\boldsymbol{\theta}_t^{(0)T} - \mathbf{m}_{<t}^{(k)}\boldsymbol{\theta}_t^{(k)T}) \tag{19}$$

$$\mathbf{a} \leftarrow \mathbf{a} - \alpha(\mathbf{m}_t^{(0)} - \mathbf{m}_t^{(k)}) \tag{20}$$

$$\mathbf{b} \leftarrow \mathbf{b} - \alpha(\boldsymbol{\theta}_t^{(0)} - \boldsymbol{\theta}_t^{(k)}) \tag{21}$$

Representation of Inter-market Coupling

As shown in Figure 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 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 3 (b) 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*. 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:

$$E(\{\boldsymbol{\theta}_{i,t}\}, \boldsymbol{\eta}_t, \{\boldsymbol{\theta}_{i,<t}\}) = -\mathbf{b}^T\boldsymbol{\eta}_t - \sum_{i=1}^{I}\mathbf{a}_i^T\boldsymbol{\theta}_{i,t} - \sum_{i=1}^{I}\boldsymbol{\theta}_{i,t}^T\mathbf{W}_i\boldsymbol{\eta}_t - \sum_{i=1}^{I}\boldsymbol{\theta}_{i,<t}^T\mathbf{A}_i\boldsymbol{\theta}_{i,t} - \sum_{i=1}^{I}\boldsymbol{\theta}_{i,<t}^T\mathbf{B}_i\boldsymbol{\eta}_t \tag{22}$$

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.

$$P(\theta_{ift} = 1 \mid \boldsymbol{\eta}_{ht}, \{\boldsymbol{\theta}_{i,<t}\}) = s(\mathbf{a}_{if} + \boldsymbol{\theta}_{i,<t}^T(\mathbf{A}_i)_{:,f} + (\mathbf{W}_i)_{f,:}\boldsymbol{\eta}_t) \tag{23}$$

$$P(\eta_{ht} = 1 \mid \{\boldsymbol{\theta}_{i,t}\}, \{\boldsymbol{\theta}_{i,<t}\}) = s(\mathbf{b}_h + \sum_{i=1}^{I}\boldsymbol{\theta}_{i,t}^T(\mathbf{W}_i)_{:,h} + \sum_{i=1}^{I}\boldsymbol{\theta}_{i,<t}^T(\mathbf{B}_i)_{:,h}) \tag{24}$$

Formalism Mapping We can find that the coupling encoding operator $\circledast$ given in Eq. (2) is implemented by term $\sum_{i=1}^{I}\boldsymbol{\theta}_{i,t}^T(\mathbf{W}_i)_{:,h}$ in Eq. (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*.

Parameter Learning The stochastic gradient update equations using CD_k can be obtained according to Eq. (13).

$$\mathbf{W}_i \leftarrow \mathbf{W}_i - \alpha(\boldsymbol{\theta}_{i,t}^{(0)}\boldsymbol{\eta}_t^{(0)T} - \boldsymbol{\theta}_{i,t}^{(k)}\boldsymbol{\eta}_t^{(k)T}) \tag{25}$$

$$\mathbf{A}_i \leftarrow \mathbf{A}_i - \alpha(\boldsymbol{\theta}_{i,<t}^{(0)}\boldsymbol{\theta}_{i,t}^{(0)T} - \boldsymbol{\theta}_{i,<t}^{(k)}\boldsymbol{\theta}_{i,t}^{(k)T}) \tag{26}$$

$$\mathbf{B}_i \leftarrow \mathbf{B}_i - \alpha(\boldsymbol{\theta}_{i,<t}^{(0)}\boldsymbol{\eta}_t^{(0)T} - \boldsymbol{\theta}_{i,<t}^{(k)}\boldsymbol{\eta}_t^{(k)T}) \tag{27}$$

$$\mathbf{a}_i \leftarrow \mathbf{a}_i - \alpha(\boldsymbol{\theta}_{i,t}^{(0)} - \boldsymbol{\theta}_{i,t}^{(k)}) \tag{28}$$

$$\mathbf{b}_i \leftarrow \mathbf{b}_i - \alpha(\boldsymbol{\eta}_t^{(0)} - \boldsymbol{\eta}_t^{(k)}) \tag{29}$$

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 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 and 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 $1 \le i \le I, T+1-2n \le t \le T$.
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 Eq. (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 $1 \le i \le I$.
$\boldsymbol{\theta}_{i,T+1} \leftarrow \boldsymbol{\theta}_{i,T}$
3. Estimate $\boldsymbol{\theta}_{i,T+1}$ for $1 \le i \le I$ by K-iteration mean-field update on second layer CCRBM, c.f. Eq. (23, 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 $1 \le i \le I$.
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 Eq. (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.

Experiment

Data Preparation

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 paper 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 1.

The data set used includes weekly closing prices from Jan 2007 to Dec 2013 [2], 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 1: Trading indexes

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

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*: 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. we use it as a baseline method.
- *Logistic*: We use this approach with indicators from the different markets in various countries, and the parameters can be obtained through MLE.
- *ANN*: We use the back-propagation algorithm in (Hyup Roh 2007) with indicators from the the different markets in various countries to train the model.

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

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

Table 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**

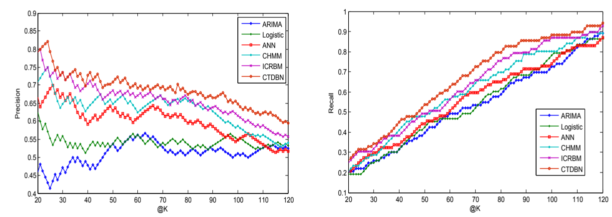
- *CHMM* (Zhong and Ghosh 2001): CHMM consists of multiple HMM chains, where each chain corresponds to model one type of financial market in a country.
- *CTDBN*: This is our deep learning approach, where the order n of the both CGRBM and CCRBM are set equal to 2 which yields good results in this experiment.
- *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.

Results

Due to space limit, 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 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 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.



(a) Precision (b) Recall

Figure 4: Precision and recall of comparative methods

Figure 4 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. 4 (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 4 (b) shows the CTDBN achieve higher recall than other models with any number of predicted trading weeks.

Conclusion and Future Applications

In this paper, 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, Hinton, and Roweis 2006), music generation (Boulanger-Lewandowski, Bengio, and 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.

Acknowledgments

This work is supported partly by Australian Research Council grants (DP130102691 and LP140100937) and National Natural Science Foundation of China (71271125).

References

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

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

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

Cao, L.; Ou, Y.; Yu, P. S.; and 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*, 85–94. ACM.

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

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

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

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

Hassan, M. R., and 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*, 192–196. IEEE.

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

Huarng, K., and 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.

Hyup Roh, T. 2007. Forecasting the volatility of stock price index. *Expert Systems with Applications* 33(4):916–922.

Jang, W.; Lee, J.; and 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.

Laitinen, E. K., and Laitinen, T. 2001. Bankruptcy prediction: application of the taylor's expansion in logistic regression. *International Review of Financial Analysis* 9(4):327–349.

Longstaff, F. A. 2010. The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics* 97(3):436 – 450.

Marcucci, J. 2005. Forecasting stock market volatility with regime-switching garch models. *Studies in Nonlinear Dynamics & Econometrics* 9(4).

Olson, D., and Mossman, C. 2003. Neural network forecasts of canadian stock returns using accounting ratios. *International Journal of Forecasting* 19(3):453–465.

Pan, H.; Tilakaratne, C.; and 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.

Taylor, G. W., and Hinton, G. E. 2009. Factored conditional restricted boltzmann machines for modeling motion style. In *Proceedings of the 26th annual international conference on machine learning*, 1025–1032. ACM.

Taylor, G. W.; Hinton, G. E.; and Roweis, S. T. 2006. Modeling human motion using binary latent variables. In *Advances in neural information processing systems*, 1345–1352.

Taylor, G. W. 2009. *Composable, distributed-state models for high-dimensional time series*. Ph.D. Dissertation, University of Toronto.

Welling, M., and Hinton, G. E. 2002. A new learning algorithm for mean field boltzmann machines. In *Artificial Neural NetworksICANN 2002*. Springer. 351–357.

Zhong, S., and 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.