Optimization of financial market risk prediction system based on computer data simulation and Markov chain Monte Carlo

Zhang Yuanfang (✉ 160072@peihua.edu.cn)
Xi’an Peihua University

Research Article

Keywords: Markov chain Monte Carlo, Data simulation, Financial market, Economic risk, Risk Assessment

Posted Date: April 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2699304/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

With the continuous development of the times, the business scope of the financial market is also expanding, and more and more unknown risks have emerged in the process of development, which has brought a certain impact to the financial market. The exchange rate fluctuation, interest rate and other factors related to the international financial market make financial market risk one of the important risks faced by securities companies. Therefore, securities companies need intuitive, accurate and effective quantitative management tools for financial market risk assessment to improve the overall risk management level of the company. Therefore, this paper constructs a financial market risk prediction system based on computer data simulation technology and Markov chain Monte Carlo algorithm. This system is mainly composed of three parts: financial risk management, data exchange management and background risk application services. The front end of the financial market risk prediction system is divided into risk monitoring, risk allocation, risk report, risk decision-making and system log modules according to functions. Data collection management functions mainly include data collection, data encryption, data conversion, data return and other functions. The measured results show that the system can predict the significant inflection points in most economic risks synchronously or in advance, and the forecast data shows that the financial security index has an upward trend from bottom to top. In this paper, the Markov chain Monte Carlo algorithm is introduced into the financial market, through data simulation to predict the risks of companies with economic risks, and make a small contribution to the market development.

1 Introduction

From 1978 to 2011, China's economic indicators continued to grow rapidly for more than 30 years, with an annual growth trend of 10.9% higher than the previous year. Since 2012, China's economic growth rate has been slower than before. The annual growth trend is 7.6% higher than that of the previous year. It can be seen from the data that China's economic development is moving from rapid growth to normalization and stability. During the period of slow economic growth, it has brought a series of changes to economic development [1]. For example, in some real economy industries, especially coal, steel and other energy and raw materials related industries. At this stage, financial institutions have undergone tremendous changes in the living environment of customers, and the asset quality of the financial industry has begun to decline [2]. In the case of slowing economic growth, an important task of the financial industry is to prevent and eliminate financial risks, especially those that may arise from overcapacity, local debt and the real estate market. On the basis of global financial development, with the continuous expansion of China's financial market, financial risks spread more rapidly and widely in the market. Moreover, at this stage, China is in the development stage of high leverage, high asset prices, high market volatility and high economic risks [3]. Under this development trend, financial management becomes more difficult. The emergence of this phenomenon has made the probability of China's financial crisis increasing, and has had a certain impact on the national and international financial development, which is also one of the major factors that have changed the country's real economy. It will lead to serious social and political
crisis, and ultimately cause irreversible impact on national financial security [4]. At present, China's economic development has entered a new stage. Because of the emergence of the financial crisis, the development of a risk estimation system has become an urgent need for China's financial management departments at this stage. In this paper, the research and development of early warning system has extremely important practical significance for China's financial security evaluation and systematic financial risk early warning [5].

2 Related Work

The literature forecasts the future trend of China's securities market through the combination method of multiple models, and establishes a continuous stochastic system for the prediction sequence [6–8]. On this basis, the dynamic prediction method of financial risk is used to quantify the market risk, and a series of studies on China's securities market risk are realized. The risk prediction of China's securities market is made according to the continuous stochastic system. With the minimum expected net loss as the dynamic financial risk measurement method, combined with the established system, the risk prediction of China's securities market is realized, and the initial investment is taken as the function of independent variables to obtain the risk prediction value [9]. Therefore, by setting different initial investments, you can create and describe two-dimensional functional images between risk and initial investment. Literature uses NET development language, Oracle database technology, the latest risk assessment and identification algorithm to design and implement a financial market risk assessment system [10]. The system provides a variety of assessment and identification models for risk, which can be used for risk measurement, and can adopt configurable strategies to adapt to the needs of different users [11]. The system has more comprehensive and mature functional modules than other financial market risk assessment systems on the market, mainly including system structure, business logic, risk monitoring, data interaction, system log, system control, risk report, risk decision-making, basic algorithms and data and other functions [12]. The literature combines normative analysis and empirical analysis to discuss the generation mechanism of financial risks [13]. Based on the assessment of China's financial security status, it puts forward relevant strategies and suggestions such as financial risks of early warning system and prevention of systemic financial risks. In the literature, data prediction is realized through the combination of multiple models on the basis of fully considering the different characteristics of two rows [14]. A continuous stochastic system is established, and the risk measurement is carried out by using the prediction sequence and dynamic financial risk measurement methods. The final results meet the expected standards. The analysis process and results can provide decision-making reference for relevant practitioners to avoid relevant deviations in the process [15].

3 Basis And Application Of Markov Chain Monte Carlo Method

3.1 Markov chain Monte Carlo principle

In the random process \(\{X_t, t = 0,1,2,\ldots\}\), for any \(t \geq 0\), and the random process may have a value in the \(\{0,1,2\}\) column or a finite number. The definition of \(X_t = i\) means that \(x\) has the state of \(i\) at time \(t\). If \(x\) has a
fixed $P_{ij}$ probability at the next time of $t + 1$ state, then the random process is called Markov chain.

Transfer probability is an important concept in Markov process, which is used to represent the conditional probability that a state at one time enters another time after a period of time. Suppose that in the process of development, something objectively has an $n$ form $E_1, E_2, \ldots, E_n$, which can only exist in one state at a time, then each state will have an $n$ form when it develops. Namely:

$$E_i \rightarrow E_1, E_i \rightarrow E_2, \ldots, E_i \rightarrow E_{i-1}, \ldots, E_i \rightarrow E_n$$

$P$ represents the probability of the existing $E_i$ state further transferring to the next $E_j$ state, then the transition probability matrix $P$:

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{pmatrix}$$

From the above equation, we can get the model $S_{t+1} = s_t * P$. That is, the forecast result is:

$$S_{t+1} = S_0 * P^{t+1} = S_0 * P_k * S^k$$

This condition can also be expressed as:

$$\pi(\theta_t) T(\theta^* | \theta_t) = \pi(\theta^*) T(\theta_t | \theta^*)$$

Obviously, the corresponding formula cannot be established without selecting an appropriate proposal distribution. To meet the conditions for fine balance, the MH algorithm can satisfy the proposed $T(\theta^* | \theta_t)$-symmetric distribution, namely:

$$T(\theta^* | \theta_t) = T(\theta_t | \theta^*)$$
In addition to symmetrical distribution design, MH algorithm also needs to determine whether new samples can be accepted:

$$A(\theta^* \mid \theta_t) = \min \left(1, \frac{\pi(\theta^*) T(\theta_t \mid \theta^*)}{\pi(\theta_t) T(\theta^* \mid \theta_t)} \right) = \min \left(1, \frac{\pi(\theta^*)}{\pi(\theta_t)} \right)$$

From the above formula, we can get:

$$\pi(\theta_t) T(\theta^* \mid \theta_t) A(\theta^* \mid \theta_t) = \min \left[ \pi(\theta_t) T(\theta^* \mid \theta_t), \pi(\theta^*) T(\theta^* \mid \theta_t) \right]$$
$$= \min \left[ \pi(\theta^*) T(\theta_t \mid \theta^*), \pi(\theta_t) T(\theta^* \mid \theta_t) \right]$$
$$= \pi(\theta^*) T(\theta_t \mid \theta^*) \min \left(1, \frac{\pi(\theta_t)}{\pi(\theta^*)} \right)$$
$$= \pi(\theta^*) T(\theta_t \mid \theta^*) A(\theta_t \mid \theta^*).$$

The transfer probability of MALA has the following forms:

$$T(\theta^* \mid \theta_n) = N(\theta^* \mid \mu(\theta_n), \epsilon^2 \cdot I)$$

$$T(\theta_n \mid \theta^*) = N(\theta_n \mid \mu(\theta^*), \epsilon^2 \cdot I)$$

$$\mu(\theta_n) = \theta_n - \frac{\epsilon^2}{2} \cdot \nabla_{\theta} \ln \pi(\theta_n)$$

$$\mu(\theta^*) = \theta^* - \frac{\epsilon^2}{2} \cdot \nabla_{\theta} \ln \pi(\theta^*)$$

MALA uses the gradient information of target distribution to build an effective Markov chain, so that the sampler can quickly converge to the target distribution.

### 3.2 Improvement of learning methods

The Bayesian network is designated as directed acyclic. All nodes in the environment are designated as $V=\{V1,..., Vn\}$. Their joint probability is as follows:
\[ P(V_1, \ldots, V_n) = \prod_{i=1}^{n} P(V_i \mid \pi_i) \]

\( \pi_i \) is a set of parent nodes \( V_i \), in which the nodes are defined by the directed acyclic structure of Bayesian networks.

\( P(V_i \mid \pi_i) \) reflects the local probability distribution among variables, which is usually reflected in the full Bayesian network through the conditional probability table. If the assumed value is discrete, the probability distribution can be determined as follows:

\[ P\left(V_i = v_j^i \mid \pi_i = \pi_k^i\right) = \theta_{ijk} \]

The main purpose of Bayesian network structure learning is to find the most likely acyclic structure after verification through experimental data. Bayesian network structure score can be defined as:

\[ P(D, G) = P(G)P(D \mid G) \]

\( P(G) \) represents the prior probability of the occurrence of the acyclic structure. \( P(D) \) represents experimental data.

MH is one of the common sampling methods in MCMC. The basic idea is that the acceptance probability of random walk corresponds to the inspection probability of the two states, which meets the requirements. It can be defined as:

\[ A = \min \left(1, \frac{P'}{P} \right) \]

In the formula, \( A \) represents the transition probability, \( P' \) and \( P \) represent the state after random walk and the probability of subsequent inspection of previous state respectively.

The purpose of independence check is to try to find out whether two nodes in the dataset meet the conditions and are independent of each other. The most commonly used statistical method is G2 statistics, which is defined as follows:
G2 statistics are subject to $\chi^2$, its degree of freedom is:

$$\text{df} = (|D(X_i)| - 1)(|D(X_j)| - 1) \prod_{X_1 \in X_k} |D(X_1)|$$

At this time, according to the actual needs, the idea of G2 statistics is used to detect, and the Max Min Parents and Children (MMPC) method is used to verify the independence of all nodes, so that an unconditional independent node set can be obtained. This method has great advantages, thus reducing the range of candidate supersets of nodes. Therefore, due to the reduction of the candidate set, the space is no longer limited, making the preprocessing method possible. Preprocessing eliminates the need to traverse and calculate the entire dataset during each evaluation, and reduces the time complexity from $O(L \times \text{node} \times 2 \text{nodes})$ to $O(\text{node} \times 2 \text{cpc})$ cpc, in terms of the number of alternative sets. Therefore, it is considered that the verification of independence is very correct. For any node, the node that is unconditionally independent and whose topology is in front of the node is the parent node.

### 3.3 Application in finance

In China's financial market model, it mainly includes bonds and stocks. For risk-free $S_0(t)$ assets, $r(t, a_i) > 0$, $i = 1, 2, \ldots$, $D$ is a bounded deterministic function, which can be regarded as the interest rate of various market models. The pricing process of risk assets $s(t)$ is as follows:

Suppose a principal grants an agent funds $x_0$ at time 0. The agent manages the investment group to manage the money within the time interval $[0, T]$, and the total fund recorded by the agent at time $t$ is $X(t)$. At each time point $t \in [0, T]$, the agent returns the funds from $c(t)X(t)$ to the customer. $C(t)$ process can be regarded as a part of agent control, and its purpose is to provide corresponding benefits for the principal from the interests of the principal.

Next, we remember that $u(t)$ is the capital invested by agent $t$ in risk assets at time $t$. We call $u(t)$ an agent's portfolio. Thus, $u(t)$ is also part of the control.

### 3.4 Simulation study

Table 1 shows the basic introduction of the multi classification problem dataset:
Table 1
Basic introduction of multi classification problem data set

<table>
<thead>
<tr>
<th>data set</th>
<th>Number of categories</th>
<th>Number of training samples</th>
<th>Number of test samples</th>
<th>Characteristic dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>usps</td>
<td>10</td>
<td>7145</td>
<td>1966</td>
<td>251</td>
</tr>
<tr>
<td>pendigits</td>
<td>10</td>
<td>7344</td>
<td>3428</td>
<td>15</td>
</tr>
<tr>
<td>covtype .binary</td>
<td>2</td>
<td>512661</td>
<td>56730</td>
<td>52</td>
</tr>
<tr>
<td>mnist</td>
<td>10</td>
<td>58800</td>
<td>9800</td>
<td>768</td>
</tr>
<tr>
<td>cifar</td>
<td>10</td>
<td>49000</td>
<td>9800</td>
<td>1003</td>
</tr>
</tbody>
</table>

Here, several simplified assumptions are made based on the basic model. We believe that the distribution probability of topic words is fixed and has parameters $\beta \in \mathbb{R}^{k \times V}$ description, where:

$$\beta_{ij} = p(w_j = 1 \mid z_i = 1)$$

Tables 2 and 3 show the posterior moments of the results of using the MCMC algorithm to study two groups of simulation data.

Table 2
A posteriori moment of the result of the GBP/USD MCMC algorithm

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\mu$</th>
<th>$\mu_q$</th>
<th>$\lambda$</th>
<th>$h\sigma$</th>
<th>$h\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True value</td>
<td>-0.177</td>
<td>0.4104</td>
<td>3.6586</td>
<td>0.0128</td>
<td>0.158</td>
</tr>
<tr>
<td>mean value</td>
<td>-0.1898</td>
<td>0.4111</td>
<td>4.2792</td>
<td>0.0124</td>
<td>0.172</td>
</tr>
<tr>
<td>S. D.</td>
<td>-0.1953</td>
<td>0.1674</td>
<td>0.9148</td>
<td>0.0003</td>
<td>0.0376</td>
</tr>
<tr>
<td>NSE</td>
<td>0.0022</td>
<td>0.00141</td>
<td>0.017</td>
<td>0.0006</td>
<td>0.0006</td>
</tr>
<tr>
<td>RNE</td>
<td>0.8439</td>
<td>1.7459</td>
<td>0.3293</td>
<td>0.2875</td>
<td>0.4673</td>
</tr>
</tbody>
</table>

The posterior moment of S&P 500 index MCMC algorithm results is shown in Table 3:
Table 3  
A posteriori moment of S&P 500 exponential MCMC algorithm results

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\mu$</th>
<th>$\mu_q$</th>
<th>$\lambda$</th>
<th>$h_\sigma$</th>
<th>$h_\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True value</td>
<td>9.8366</td>
<td>-0.178</td>
<td>3.1106</td>
<td>0.005</td>
<td>0.0553</td>
</tr>
<tr>
<td>mean value</td>
<td>10.685</td>
<td>-0.199</td>
<td>2.8371</td>
<td>0.0048</td>
<td>0.0595</td>
</tr>
<tr>
<td>S. D.</td>
<td>2.328</td>
<td>0.1903</td>
<td>0.742</td>
<td>0.0001</td>
<td>0.016</td>
</tr>
<tr>
<td>NSE</td>
<td>0.0321</td>
<td>0.0025</td>
<td>0.011</td>
<td>0.0001</td>
<td>0.0003</td>
</tr>
<tr>
<td>RNE</td>
<td>0.6469</td>
<td>0.6692</td>
<td>0.4887</td>
<td>0.5005</td>
<td>0.44</td>
</tr>
</tbody>
</table>

With the further development of the reconfigurable computing system composed of random gates, the system composed of random gates can sample any graphical model. In order to make the random gates interconnect and conform to the arbitrary graphic structure, we adopt a reconfigurable interconnection method in the random gate interconnection structure to make any two random gates interconnect. In order to realize different graphical models, it is also necessary to configure the conditional realization probability of each random gate. Therefore, we take the previously developed reconfigurable random gate as the basic module of the system.

See Fig. 1 for details.

It can be seen from Fig. 1 that the increase in the number of nodes will lead to the change of not introducing pipeline, while the implementation of sampling algorithm with pipeline is basically fixed in execution time, which is not affected by the increase in the number of nodes. In contrast, as shown in Fig. 1, the time required to execute a graphics model based sampling algorithm in a traditional processor is also measured. Compared with the graphics model with 64 nodes, the implementation speed based on random gates without pipeline is almost 30 times that of traditional processors, while the implementation speed based on random gates with pipeline is almost 2000 times that of traditional processors. The results verify the performance advantages of the random gate system, and prove the feasibility of its accelerated sampling algorithm based on graphical model.

As shown in Table 4:
<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Accuracy of general methods</th>
<th>Accuracy of improvement method</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>20</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>25</td>
<td>0.93</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Through the comparison of the methods in this paper, it is found that better accuracy can be obtained. If only the correct rate of finding edges is calculated, it usually reaches more than 95%, but usually only less than 5% of the wrong edges will be added. Compared with the general sampling method, the loss of accuracy is not obvious.

4 Optimization Of Financial Market Risk Prediction System And Computer Data Simulation Analysis

4.1 System logic architecture

The financial market risk prediction system consists of three parts: financial risk management, data exchange management and background risk application services. Figure 2 is the system logic diagram:

4.2 Processing flow of financial data system

The background process flow chart of the financial market risk estimation system is shown in Fig. 3.

4.3 Data simulation experiment results and analysis

Figure 4 shows the multi-stage prediction of the comprehensive financial security index from 2006 to 2021 achieved through the in-depth learning model. The blue is the real value and the red is the predicted value.

As can be seen from Fig. 4, almost every major inflection point can be synchronized or predicted in advance. The forecast data shows that the financial security index is on the rise from bottom to top.

After the initially defined variables are reduced, the model is basically established. However, in the initial frequency determination, there may be a problem of missing variables, leading to some significant variables and some frequencies not being considered by the model, and the final prediction results will have a large deviation. In order to avoid this problem, the model residuals must be checked as necessary. Through calculation and analysis, the residual time sequence diagram is shown in Fig. 5.
5 Research On Early Warning And Prevention And Control Strategies Of Financial Market Risks

5.1 Policy Suggestions on Preventing Systemic Financial Risks

In order to increase financial support for the real economy, the following specific measures must be taken:

(1) Change the profit model of commercial banks. By increasing equity, debt markets and other channels, we can increase our own capital, improve the capital adequacy ratio, and further improve the ability and level of independent and reasonable pricing and risk prevention of enterprises. Encourage commercial banks to manage the cost of debt scientifically, increase the proportion of intermediary business in the total business, improve the capital adequacy ratio, and reduce the dependence on interest margin income.

(2) Strengthen training on financial risks, and timely investigate and resolve various financial risks. Continue to strengthen the follow-up risk analysis of macroeconomic situation and import and export trade, timely track and pay attention to changes in the asset quality and liquidity of commercial banks, and constantly adjust financial risk response measures.

5.2 Building a financial risk early warning system suitable for China's national conditions

At present, China lacks a relatively complete financial database covering the whole country. Establishing such a database as a collection channel for data exchange can reduce the cost of data collection and develop data standards to facilitate data processing and improve the reliability and uniformity of data. At the same time, such a large financial database is essential to improve China's international financial competitiveness and is an important symbol of China's ability to cope with financial risks. At present, China is in urgent need of a large financial database, which also has a good foundation. Although there is no direct establishment of a financial database, banks and other financial institutions have consciously stored financial business data for their own needs, which effectively improves the development trend of financial databases. At present, there is a certain gap between China's financial supervision level and that of developed countries, and some regulatory infrastructure needs to be improved. For example, China has not yet established a large-scale financial database covering the entire financial field across industries. The current proposal is to let the People's Bank of China take the main responsibility to achieve the data integration of insurance, securities and banking. At the same time, in view of the different needs of these three industries, it is also necessary to establish targeted sub databases suitable for these three sectors.

6. Conclusion
The process of financial globalization has promoted the close connection of China's internal financial markets and greatly increased the probability of financial market risks. At this stage, China is in the development stage of high leverage, high asset prices, high market volatility and high economic risks. Under this development trend, financial management has become more difficult, and the probability of financial crisis in China has been increasing. China is moving from rapid growth to normalization and stability in the field of economic development. The economic downturn, overcapacity and the rise of Internet finance have changed the business environment of commercial banks. The new normal of economy is pushing the development of commercial banks and even the capital market into a new stage, which may be accompanied by many adverse factors. At this stage, financial risks will become increasingly prominent, which has great practical significance for the study of financial security issues. Therefore, this paper establishes a financial market risk prediction system based on computer data simulation technology and Markov chain Monte Carlo algorithm.

**Declarations**

**Conflict of interest**

The authors declare that they have no conflict of interests

**Ethical approval**

This article does not contain any studies with human participants performed by any of the authors.

**Data Availability**

Data will be made available on request.

**References**


Figures
Figure 1

Performance comparison between Bayesian network based sampling and traditional processor based sampling algorithm using a reconfigurable computing system composed of random gates.
Figure 2

System logic architecture diagram
Figure 3

Background Process Flow Chart of Financial Market Risk Estimation System
Figure 4

Comparison Chart of Single step Prediction of 2006-2011 Financial Security Index

Figure 5

Residual Sequence Sequence Diagram of Shanghai Composite Index Model