Journal of Organizational and End User Computing
Volume 36 • Issue 1

Financial Cycle With Text Information Embedding Based on LDA Measurement and Nowcasting

Peijin Li, Shanghai University of International Business and Economics, China*
https://orcid.org/0000-0003-0248-999X

Xinyi Peng, East China Normal University, China

Chonghui Zhang, Zhejiang Gongshang University, China

Tomas Baležentis, Lithuanian Institute of Agrarian Economics, Lithuania
https://orcid.org/0000-0002-3906-1711

ABSTRACT

When compared to traditional indicators, text information can capture market sentiment, investor confidence, and public opinion more effectively. Meanwhile, the mixed-frequency dynamic factor model (MF-DFM) can capture current changes. In this study, the authors constructed a financial cycle measurement and nowcasting framework by incorporating text information into factors derived from MF-DFM. The findings reveal that, first, the financial cycle indicator (FCI) provides a more detailed and forward-looking perspective on major events. Second, it can serve as an effective "early warning system" by cross-referencing economic indicators. Third, financial cycles exhibit five short cycles, with contraction periods being longer than expansion phases and expansion amplitudes surpassing contractions. Lastly, the analysis suggests a potential turning point in the second half of 2023. This research represents a valuable attempt to integrate big data for more sensitive, timely, and accurate monitoring of financial dynamics.

KEYWORDS

Financial Cycles, LDA, Mixed-Frequency Dynamic Factor Model, Sentiment Analysis, Text Mining

1. INTRODUCTION

During the long-term development, financial cycles are considered as macro-financial conditions like credit scale, capital flow, leverage level, and asset prices such as stocks, bonds, exchange rate, and derivatives, that have been experienced occasionally extreme booms and busts. More accurate identification, definition, and monitoring of the financial cycle can help enhance risk awareness and promote macroeconomic stability, thereby effectively preventing risks. In terms of financial cycle measurement, Dynamic Factor Model (DFM), whose principle is to reduce the dimensionality

DOI: 10.4018/JOEUC.335082
*Corresponding Author

of the data, use one or more extracted dynamic factors to explain the fluctuations, and the factors themselves can also be used for Nowcasting and can predict situations for recent past, current, and short-term future.

Most indicators from financial markets or macroeconomic indicators have limited abilities to capture the full picture of financial activities especially for non-traditional sectors. The emergence of new types of data represented by text information such as news articles and social media posts has opened up new avenues for research. On the one hand, text information can provide governments, central banks, and markets with more timely information on short-term prospects, potential risks and vulnerabilities. On the other hand, it can provide economic sentiment and behavioral insights that traditional indicators cannot provide.

This article aims to construct a methodology on embedding text information into financial cycle measurement and Nowcasting, to obtain more accurate and timely financial cycles. Specifically, by text mining financial news from 2002 to 2022, 10 topics are constructed and their corresponding topic probability distribution form text information sequences based on the Latent Dirichlet Allocation (LDA). Furthermore, the above information sequence combined with 37 traditional indicators from the comprehensive financial risk indicator system form 47 sequences. By building a Mixed Frequency Dynamic Factor Model (MF-DFM) and extract common factors as a new financial cycle index for measuring cycles and conduct Nowcasting.

The potential marginal contribution may include the following. Firstly, enriching the financial cycle measurement. By embedding text information that can include the emotions, public opinions, events and other factors of market participants, the research can not only be more comprehensive, but also conducive to in-depth interpretation of the financial market. Secondly, expanding the application of the big data on finance. Unlike TF-IDF, the LDA can timely integrate financial fluctuations concerning sentiment that can support investment decision-making, risk management, and policy formulation. Thirdly, timely monitoring and early warning the development of the financial cycle. Real-time Nowcasting for financial markets can help capturing changes in market sentiment and public opinion in advance, which is of great significance to investors, regulatory agencies, and policy makers.

2. LITERATURE REVIEW

2.1 Financial Cycle Indicators

As a barometer of financial prosperity, financial cycle is an important indicator to accurately determine the cyclical fluctuation of the financial system, to in-depth explore the linkage between finance and macroeconomics, and to implement regulatory (Deng et al., 2022). However, differences exist in both indicators and methodologies. Early research focused on credit cycle by the amount of credit or its ratio to GDP (Schularick & Taylor, 2012; DeBonis & Silvestrini, 2014; Aikman et al., 2015). Although the single indicator is simple, it also overlooks the potential interacts between credit and other financial assets. Multi indicators as Monetary Condition Index (MCI) and Financial Conditions Index (FCI) incorporates interest rate, exchange rate, stock price, credit and real estate price that can reflect the future financial activities (Hansson et al.,1994; Goodhart & Hofmann, 2001; Lack, 2003; Hatzius et al., 2010; Drehmann et al., 2012). Indicators of balance sheet structure like non-core liabilities to broad money ratio (nc) and so on can also be used (Krupkina & Ponomarenko, 2015). Although research focus on the multi indicators, the overall logic is essentially consistent, that is, the financial cycle mainly reflects the “financing constraints of the financial market” and “risk and asset value”. However, research presents relatively insufficient availability and comparability.

2.2 Financial Cycle Measurement

In terms of the financial cycles, many methodologies are extensively used. The Bry-Boschan Turning Point Analysis (B-B method) is mostly chosen by scholars (Bry & Boschan, 1971). Claessens et al.

(2011) analyze the interaction between financial cycle and business cycle. The HP Filter or BP Filter are also widely used (Drehmann et al., 2012; Aikman et al., 2015). However, disadvantages of the limitations in theory mainly oriented on implementing certain principles in advance. Strohsal et al. (2015) applied a parametric method to estimate the spectral densities to measure the main lengths of financial cycles. Pontines (2017) used Spectral Analysis to measure the cycles of four East Asian economies, which can better capture the characteristics of financial cycles.

DFM has unique advantages for the extension of the classical factor model on time series (Geweke et al., 1977). MIDAS (Mixed Data Simpling) can model data of different frequencies and maximize the use of information from different frequencies (Ghysel, 2004). Arunba (2009) constructed a state-space model with mixed characteristics through the concept of stock and flow, and then constructed a real-time condition index through variables covering daily, weekly, monthly and quarterly frequency. Banbura & Modugno (2014) considered more variables and integrated the mixing state-space model in the case of large variables.

2.3 Text Mining in Finance

In terms of finance, text information is mainly applied to five indicators as readability, attention, emotion, implied volatility and disagreement (Yan et al., 2019). Gunning (1952) first proposed the Fog Index to measure the readability difficulty. Da et al. (2011) expressed investors' attention by measuring the number of searches conducted on Google. Soo (2018) scored articles from multiple news media by sentiment analysis, and proved that it can be an important factor to affect the changes of the real estate market. LDA can help to automatically extract valuable information from text data. After Deer (1990) proposed Latent Semantic Analysis (LSA), Hofmann (1991) proposed the Probabilistic Latent Semantic Analysis (PLSA), Blei (2013) proposed the LDA, which incorporates prior distribution into PLSA. Ryohei et al. (2013) extracted corresponding topics from 24 million news records. Thoursrud (2019) used the LDA to extract 80 topics from business newspapers, an constructed index using the MF-DFM. Text information provides insights into economic sentiment and behavior that traditional data cannot, thus can greatly expand the scope of financial research.

3. METHODOLOGY

3.1 LDA

3.1.1 Description of Parameters

LDA is constructed for automatically identifying topics and their distributions in large text corpora. The Dirichlet distribution is the conjugate prior distribution of a multinomial distribution, and its density function as follows:

$$f\left(p_1,\ldots,p_{k-1};\alpha_1,\ldots,\alpha_{k-1}\right)=\frac{1}{\Delta\left(\vec{a}\right)}\prod_{i=1}^{k}P_i^{\alpha_i-1} \tag{1}$$

$$\Delta\left(\vec{a}\right)=\frac{\prod_{i=1}^{k}T\left(\alpha_i\right)}{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)} \tag{2}$$

3.1.2 Gibbs Sampling

As a Markov Chain Monte Carlo (MCMC) algorithm, Gibbs Sampling is suitable for sampling from the probability distributions.

- Randomly select the word w from the m-th document d_m to randomly assign a topic Z.

Table 1. Mathematical Notation in LDA

Notation	Meaning
M	Number of documents
D	Document Collection
V	Number of words in the document set
K	Number of topics
$w_{i,j}$	A specific word
$Z_{i,j}$	The j-th theme word of the i-th document
α	Prior Dirichlet distribution parameters for the topic distribution of each document
β	Prior Dirichlet distribution parameters for each topic word distribution
N_i	Length of each document (document i contains N_i words)
θ_i	The Topic distribution of document i (determined by distribution of α)
ϕ_k	The word distribution of theme k (determined by the distribution of β)

- Calculate $P(w_i \mid Z_i)$, the probability of words appearing under each topic Z_i, $P(w_i \mid Z_i, d_m)$ represents the probability of words appearing under each topic Z_i in each document m, and the probability distribution of each word.
- Recreate a new topic and update the topic and probability distribution through multiple iterations to obtain the topic distribution in each document and the word distribution under each topic. Repeat the process until the topic distribution and the word distribution converge.

3.1.3 Model Structure

As shown in Figure 1, LDA is a Bayesian model consisting of three layers, documents, topics and words. Documents and words are observable, while topics are hidden. The three layers are connected by probability distributions, where the thickness of the line indicates the magnitude of the probability. The higher the probability, the thicker the line, and vice versa. The basic idea of the model is to select a topic based on a certain probability and select a word from that topic.

As shown in Figure 2, each document can be considered as a combination of multiple topics, and the probability distribution of the topics reflects the distribution of documents over all topics. At the same time, each topic is also a probability distribution of words, indicating the importance of each word under that topic. Therefore, documents and words can be analyzed by projecting them both into the same latent semantic space. The topic space, thus revealing the intrinsic connections and underlying meanings between documents and words.

The LDA is generated in the following steps:

Figure 1. Three-level structure of the theme model

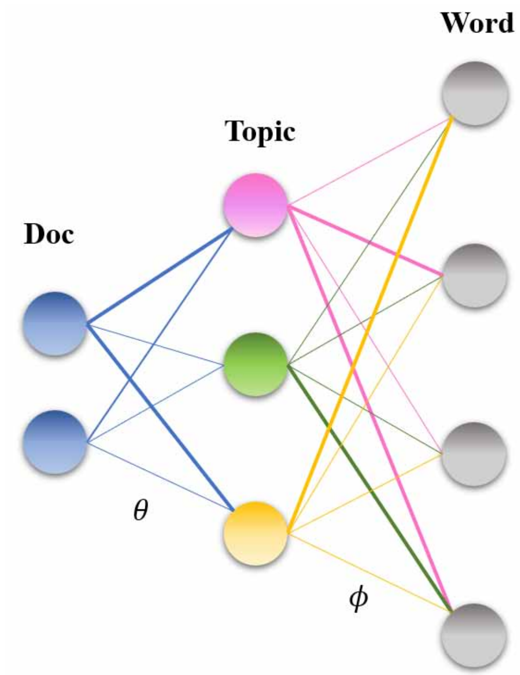
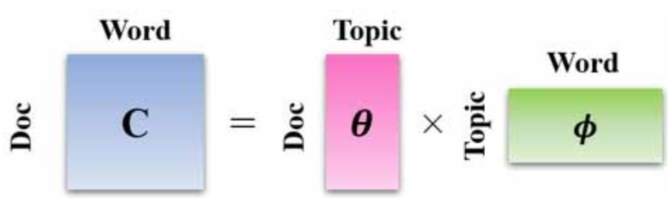


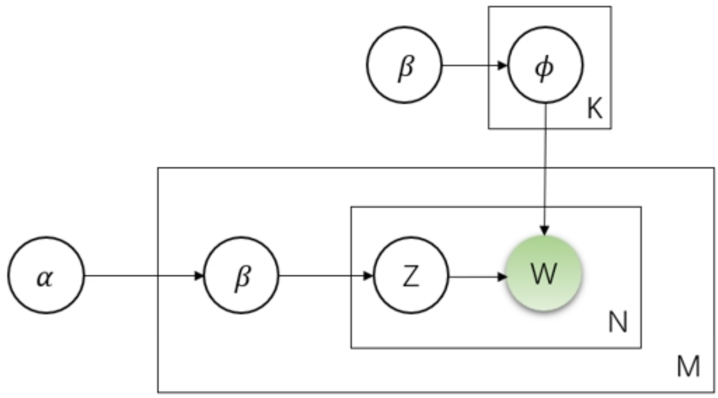
Figure 2. Matrix decomposition of LDA



- Assuming there is a total of M documents, each generated by a mixture of multiple topics, and the length of the first document is N_i(i=1, 2, …, M).
- Select θ_m, the topic distribution of the document (following the Dirichlet distribution of α), m =1,2, …, M.
- Obtain a topic Z from the topic distribution θ_m.
- Select a word distribution ϕ_k (following the Dirichlet distribution of β, k =1, 2, …, K) for the topic K, with the number of K pre-set.
- Obtain a word w_n from word distribution ϕ_k, which is the n-th word in the m-th document.

As can be seen in Figure 3, the model contains hidden variables and observable variables that are conditionally dependent on each other (the arrow A pointing to B indicates that B is determined by A). The matrix box indicates repeated sampling. The letters in the lower right-hand corner of the matrix indicate the number of repeated samples.

Figure 3. The training process of LDA



Based on the above document generation process, the joint distribution probability of each word w, the topic Z to which each word belongs, and the topic distribution θ of each document can be obtained.

$$p\left(w,z,\theta|\alpha,\beta\right)=\prod_{j=1}^{N}p\left(\theta_i|\alpha\right)p\left(z_{i,j}|\theta_i\right)p\left(\varphi|\beta\right)p\left(w_{i,j}|\theta_{z_{i,j}}\right) \quad (3)$$

The training process of the LDA model consists of the following steps:

- Randomly initialize and assign a topic $k\in K$ to each word $w\in V$ in the corpus.
- For each word w, the Gibbs sampling formula is used to resample its topic and update it in the corpus.
- Repeat previous step until the Gibbs sampling convergence no longer changes.

3.1.4 Conversion of Structured Data

The LDA model is trained to obtain the probability distribution of the lexical items in each topic and the probability distribution of the topics. For an LDA containing K topics, it is assumed that the probability of a document d containing topics K is $\theta_{d,k}$. By averaging the $\theta_{d,k}$ of each topic over a fixed time period, the strength of document d can be obtained as follows:

$$strength_{t,k}=\frac{\sum_{i=1}^{n_t}\theta_{d_i,k}}{n_t} \quad (4)$$

Where $strength_{t,k}$ denotes the theme intensity of the k-th theme under the time t, $\sum_{i=1}^{n_t}\theta_{d_i,k}$ denotes the sum of probabilities corresponding to all k topics at time t, n_t denotes the number of documents corresponding to the length of time.

3.1.5 Sentiment Analysis

Three types of methods for sentiment analysis, which are dictionary, machine learning, and deep learning.

a. Identifying the Sentiment Dictionary

The Chinese basic emotion dictionaries in common use include HowNet, National Taiwan University Sentiment Dictionary (NTUSD), and the ontology library of Dalian University of Technology Chinese Sentiment Dictionary. English basic sentiment Dictionaries include WordNet and SentiWordNet. Assign certain full weights to positive and negative emotional vocabulary respectively.

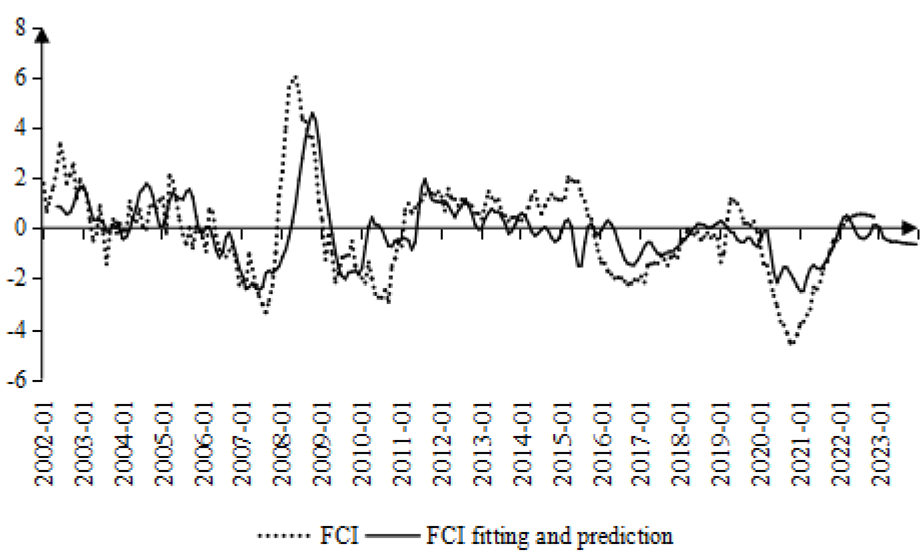
b. Emotional Vocabulary for Matching Participles

Iterate through the obtained word separation results, locate the sentiment word according to the sentiment dictionary object, the initial sentiment word weight is assigned as (w). If there is a negative word, the next sentiment word weight value is reversed (-w). If there are multiple negatives, the sentiment is determined according to the number of negatives in the previous text.

c. Calculating the Sentimental Scores

The sentiment score is presented as *Weight* × *Initial*. The sentiment score of each word is added to through the whole text. Repeat the above steps until have traversed the entire list of words, and the sentiment score for the whole text is obtained. The score, which is the sum of the sentiment scores of all the sentiment words, can be used to represent the sentiment tendency and intensity of the text.

Figure 4. General process of sentiment analysis based on dictionaries



3.2 Dynamic Factors

Economic research is faced with the insufficient data frequency. DFM is applicable to reduce the dimensionality by integrating high-dimensional data into a few factors. Moreover, it can also capture the dynamic characteristics and predict macroeconomic conditions.

3.3.1 The Equation and Parameter

As a type of the state space equation, the DFM assumes that observable data is driven by a low dimensional unobservable process f_t. The model expressed as follows:

$$X_t = C(L)f_t + \varepsilon_t \tag{5}$$
$$f_t = A(L)f_{t-1} + \mu_t \tag{6}$$

Where X_t contains N sequences, is a vector of observables that may contain missing data. X_t, ε_t are $N \times 1$ matrix. The number of dynamic factors is q and f_t is a $q \times 1$ matrix while L is the lag operatoe. $C(L)$ and $A(L)$ are respectively $N \times q$ and $q \times q$ matrix. The i-th line $C_i(L)$ is the loading coefficient of the i-th sequence $X_{i,t}$ of X_t. For any k, there is $E\varepsilon_t \mu'_{t-k} = 0$, and $E\varepsilon_{it}\varepsilon_{js} = 0, i \neq j$.

Following the p-order autoregressive process, we obtain:

$$f_t = A_1 f_{t-1} + \ldots + A_p f_{t-p} + \mu_t \tag{7}$$

Suppose that $A(L)f_t = \mu_t$, where $\mu_t \sim N(0, I_r)$, $A(L)$ is a lagged polynomial with roots in the unit circle, which means that f_t is the covariance-smooth process.

Assuming the existence of q factors, namely q dimensional unobservable process f_t and q dimensional observable process μ_t. If f_t is observed, OLS can be used to estimate { $A_1, \ldots A_p$ } in the context of the $VAQ(p)$. The estimator is also a maximum likelihood estimator.

Let $f_t = (f_{1,t}, \ldots, f_{q,t})$, $\mu_t = (\mu_{1,t}, \ldots, \mu_{q,t})$, then the equation is:

$$\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ \cdots \\ f_{q,t} \\ f_{1,t-1} \\ \cdots \\ f_{q,t-p+1} \end{pmatrix} = \begin{bmatrix} A_1 & \cdots & \cdots & A_p \\ I_q & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & I_q & 0 \end{bmatrix} \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ \cdots \\ f_{r,t-1} \\ f_{1,t-2} \\ \cdots \\ f_{q,t-p} \end{pmatrix} + \begin{pmatrix} \mu_{1,t} \\ \cdots \\ \mu_{q,t} \\ 0 \\ \cdots \\ 0 \end{pmatrix} \tag{8}$$

The three main types of model estimation are PCA estimation, two-step estimation and QML estimation. The QML estimation use PCA estimation and the Kalman filters to iterate until Expectation Maximization (EM) converges. Observations are used to estimate the parameters in a maximum likelihood, while considering the presence of missing data.

3.3.2 Mixed Frequency

For x_t to represent a potentially common factor that is unobservable at period t, the unobservable factor can generally be assumed to be subject to an $AR(p)$ process with covariance stability as follows:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \ldots + \rho_p x_{t-p} + e_t \tag{9}$$

Where, the random error term $e_t \sim N\left(0, \sigma_e^2\right)$, and $\left(\rho_1, \rho_2, \ldots, \rho_p, \sigma_e^2\right)$ should be $\sigma_x^2 = 1$, that is, the variance of the unobservable factor is 1. Using the method of Aruoba et al. (2009) for reference, the summation operator C_t is introduced into the factor model to deal with the summation problem of unobservable variables. Here, the quarterly summation operator C_t is defined as the quarterly summation of the monthly unobserved factor x_t, expressed as:

$$C_t = \xi_t C_{t-1} + x_t = \xi_t C_{t-1} + \rho_1 x_{t-1} + \rho_2 x_{t-2} + \ldots + \rho_p x_{t-p} + e_t \tag{10}$$

Among them

$$\text{¾} = \begin{cases} 0, & T\ is\ the\ first\ month\ of\ the\ quarter \\ 1, & others \end{cases} \tag{11}$$

The observable variables are denoted as $\hat{y}_t^i$. These observable and unobservable factors have the following linear relationship. For the i^{th} monthly variable $\hat{y}_t^i$, it can be expressed as the common factor and the heterogeneity component:

$$\hat{y}_t^i = c^i + \beta^i x_i + \varepsilon_t^i \tag{12}$$

where c^i is a vector of constants, β^i is a vector of coefficients, and the heterogeneity component ε_t^i obeys the $AR(p)$ process.

$$\varepsilon_t^i = \gamma_1^i \varepsilon_{t-1}^i + \gamma_2^i \varepsilon_{t-2}^i + \ldots + \gamma_p^i \varepsilon_{t-p}^i + v_t^i \quad v_t^i \sim N\left(0, \sigma_{vi}^2\right) \tag{13}$$

3.3.3 The Number of Factors

Bai & Ng (2002) proposed a method to consistently estimate the optimal number of common factors under the assumption that there is a large cross section n and a small time dimension T, that is, by setting some penalty criteria to determine the optimal number q of common factors. For different penalty functions, the following three information criteria are presenteds:

$$IC_{p1}(q) = \ln V_q\left(\hat{\Lambda}, \hat{F}\right) + q\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right) \tag{14}$$

$$IC_{p2}\left(q\right) = \ln V_q\left(\hat{\Lambda}, \hat{F}\right) + q\left(\frac{N+T}{NT}\right)\ln\left(C_{NT}^2\right) \tag{15}$$

$$IC_{p3}\left(q\right) = \ln V_q\left(\hat{\Lambda}, \hat{F}\right) + r\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) \tag{16}$$

$C_{NT} = \min\left\{\sqrt{N}\sqrt{T}\right\}$. According to the criteria, the number of common factors q is chosen to minimise the above three criteria. It should be noted that the number of public factors determined by this criterion is only an upper limit to the true number of dynamic factors.

3.3 Nowcasting

According to Giannone et al. (2008), DFM can be applied to Nowcasting, by inheriting the idea of DFM and achieving an increase in the frequency of prediction variables. The observable variable X_t is represented as a linear combination of an unobservable common factor F_t and a trait disturbance term e_t, where F_t only includes the current period value and does not include the lag term, that is, $F_t \sim VAR\left(1\right)$. The specific expression is as follows:

$$X_t = \Lambda F_t + e_t \tag{17}$$

$$F_t = AF_{t-1} + B\eta_t \tag{18}$$

$F_t = \left(f_{1t}, f_{2t}, \dots f_{rt}\right)$ is the vector composed of r common factors. Λ is the $n \times r$ factor loading matrix. A is the $r \times r$ matrix and B is the $r \times q$ matrix. The covariance matrix Σ_e is orthogonal. Nowcasting also requires combining the monthly DFM with a mean-adjusted quarterly series forecasting equation. The potential monthly series can be expressed as a linear function of common factors:

$$y_t^M = \beta' F_t + \varepsilon_t, \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \tag{19}$$

The Nowcasting for the quarterly series are:

$$\hat{y}_{3k}^Q = \frac{1}{3}\beta'\left(\hat{F}_{3k} + \hat{F}_{3k-1} + \hat{F}_{3k-2}\right) = \beta'\hat{F}_{3k}^Q \tag{20}$$

$\left(X_{1:T}, \hat{A}, \hat{C}\right)$ can be estimated based on the results of the dynamic factor. For $h \in \left\{1,2,\dots\right\}$, the prediction $\left(X_{f,T+h}, f_{T+h}\right)$ can be calculated in the following three steps:

First, given the estimate of f_T and $VAR\left(p\right)$, $f_{f,T+1}$ can be predicted as:

$$f_{f,T+1} = \hat{A}_1 f_T + \dots + \hat{A}_p f_{T-p} \tag{21}$$

Next, $X_{f,T+1}$ can be predicted based on the factor loadings:

$$X_{f,T+1} = \hat{C} f_{f,T+1} \tag{22}$$

Finally, repeat step1 and step2 until the desired level is reached.

4. RESULTS

4.1 Text Information in LDA

4.1.1 Data Collection and Preprocessing

We extract the sequences by LDA to add text information to the financial cycle. 46719 news reports from *China News Network* with titles containing "Finance" were obtained from January 1, 2002, to December 31, 2022. The data was pre-processed to remove 9421 reports of similar, job postings and other irrelevant texts, resulting in 37,298 pieces of text data. Word segmentation, custom vocabulary, and dictionary of inactive words establishing are executed.

Before constructing the LDA, two hyperparameter α and β need to be preset. α represents the probability distribution between the document and the topic, and its value is related to the number of topics k. The larger α, the more evenly distributed the theme of the document. β represents the probability distribution between topics and words, and its value is related to the number of words w in the corpus. A larger β indicates that there are more word items in the document. As a rule of thumb, $\alpha = 50 / k$, $\beta = 0.01$. In the determination of the number of topics K, we use two indicators of Topic Perplexity and Topic Consistency.

Topic Perplexity. Perplexity is inversely proportional to the prediction ability of the model. Put the unknown documents into the topic model, and then compare the inferred results with the original documents to calculate the quality of prediction. The equation of the Topic Perplexity is:

$$p\left(w_m\right) = \prod_{i=1}^{N_\alpha} \sum_{Z} p(w_{d,i}|z)p(z \mid d) \tag{23}$$

$$Perplexity = exp\left\{-\frac{\sum_{m=1}^{M} \log p\left(w_m\right)}{\sum_{m=1}^{M} N_m}\right\} \tag{24}$$

$p\left(w_m\right)$ denotes the probability of generating a document m, N_m denotes the number of words in the first m the number of words. The lower the perplexity, the better the predictive ability of the model.

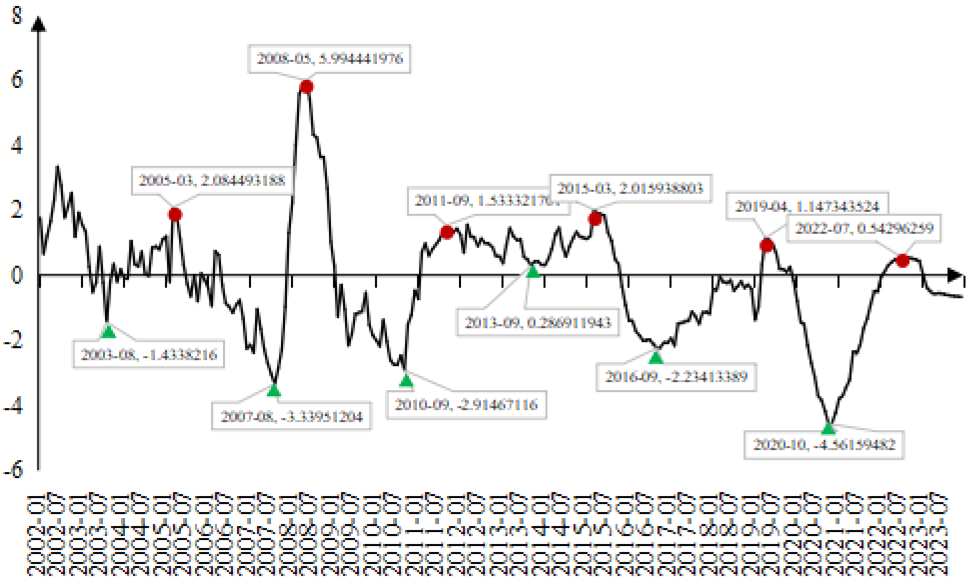
Topic Consistency. Consistency is to measure the interpretability of the topic. If it is easy to interpret, the top word of the topic should appear more frequently in the corresponding corpus, thereby improving consistency. The equation of the Topic Consistency is:

$$C_k = \sum_{m=2}^{M} \sum_{l=1}^{m} log \frac{D\left(w_m^k, w_l^k\right) + 1}{D\left(w_l^k\right)} \tag{25}$$

$W^k = \left(w_1, \ldots, w_m^k\right)$ is a list of M top words in topic k, $D\left(w\right)$ is the number of financial reports containing words w, and $D\left(w, w'\right)$ is the number of financial texts where words w and w' have appeared together at least once. The higher the consistency, the better the predictive ability.

As shown in Figure 5, the number of topics of the LDA is traversed from 1 to 20. According to Topic Perplexity, when the number of topics increases from 1 to 6, the Perplexity has a clear downward

Figure 5. Perplexity (left), consistency (right)



trend. Perplexity ushered in an obvious turning point after 6, and the decrease trend of perplexity slowed down. According to Topic Consistency, when the number of themes is more than 10, the consistency score has reached a high level. Therefore, the theme number $K = 10$ is determined and the LDA topic model established.

4.1.3 Theme Words and Distribution

4.1.3.1 Topic Clustering

Setting $\alpha = \dfrac{50}{k}$, $\beta = 0.01$, $k = 10$, the number of iterations as 50. LDA is trained using the batch learning method to obtain the topic lexical item probability distribution and the topic probability distribution. The four quadrant diagram shown that the topic clustering works well.

As shown in Table 2, topics 1-10 are defined as stocks, policies, regional cooperation, credit, market order, globalization, technology, investment, banking, and social financing.

4.1.3.2 Document-Topic Distribution

For LDA with 10 topics, a document d contains the topic k with a probability of $\theta_{d,k}$, as shown in Table 3.

Figure 6. Four quadrant diagram of LDA topic model

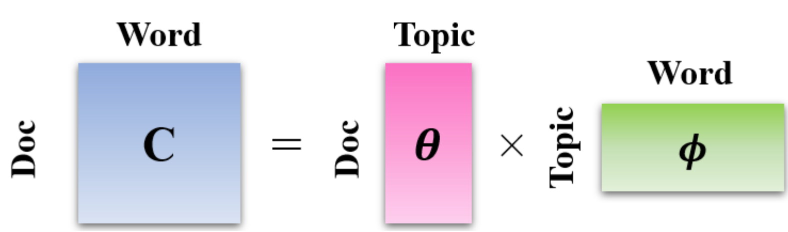


Table 2. Main theme words for different topics

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Stock	Policy	Regional Cooperation	Credit	Market Order	Globalization	Sci. & Tech	Investment	Bank	Social Finance
ponents	supervisory	collaborative	interest rates	consumers	global	the Internet	company	bank	enterprise
ctonic	reforming	global	CNY	work	financial crisis	platform	invest	business	loan
profession	Risk	construct	loan	information	international	industry	fund	custom	financing
security	Financial institution	CNY	financing	activity	car	consumption	property	green	financial service
Stock market	capital	financial center	deposit	case	crisis	Sci. & Tech	business	service	service
increase	system	enterprise	currency	investigate	country	data	group	banking	rural area
decline	institution	industry	fund	risk	effect	service	manage	product	financial institution
fund	folk	project	bond	product	world	technology	security	account	credit
bank	work	invest	financial institution	fund	financial market	product	enterprise	financial institution	fund
anticipate	Financial risk	center	bank	company	government	enterprise	industry	financial service	policy

Table 3. Document-topic probability distribution

Text	Time	Stock	Policy	Regional Coop	Credit	Market Order	Globalization	Sci. & Tech	Investment	Bank	Social Finance
China will continue to implement financial policies to support the development of Tibet……	2002/1	0.0017	0.3467	0.0411	0.1509	0.0017	0.0017	0.0017	0.0017	0.0663	0.3865
Experts believe that cooperation between banks and securities is an upgrade of China's financial capabilities……	2002/1	0.0011	0.7264	0.1731	0.0680	0.0011	0.0256	0.0011	0.0011	0.0011	0.0011
Profits declined significantly due to Economy of Argentina's economic crisis……	2002/1	0.0029	0.0029	0.0029	0.4419	0.0029	0.3052	0.0029	0.1298	0.1055	0.0029
……	…	…	…	…	…	…	…	…	…	…	…
The financial market has not entered the "once in a decade" fluctuation cycle……	2022/12	0.1668	0.0008	0.0008	0.3083	0.0356	0.3970	0.0008	0.0008	0.0448	0.0445

4.1.3.3 Sentiment Analysis

Beyond HowNet, NTUSD, we also consider the words according to financial theories. Each positive word is assigned a weight of 1, and negative word assigned of -1. The sentiment values were assumed to satisfy the linear superposition principle. Due to the existence of multiple negations in Chinese, it indicates a negative meaning. When a negative word appears even times, it indicates a positive meaning. Calculate the emotional score of each text, and when the score ≥ 0, the text conveys positive. When the score < 0, the text is judged to be negative. The probability values in Table 3 were processed positively or negatively according to the emotions, as shown in Table 4.

The strength of a topic is calculated based on the probability that each document corresponds to a topic. The strength of a topic refers to its attention in a time window and is proportional to the probability of the topic being included in a document. In order to extract the theme intensity of 10 themes into monthly sequence data, we calculated the monthly mean of document topic probability for each month, and obtained the monthly theme intensity as shown in Table 5. The sequence visualization is shown in Figure 7.

It can be observed that the topics of credit, globalization, investment, banking, and market order have significant fluctuations in the first 10 years, while for the latter 10 years remained relatively flat. The popularity of technology topics has significantly increased in the past decade. The peak popularity of stocks was mainly concentrated from 2006 to 2015. The news popularity of social financing is mainly concentrated in 2020-2022. Thus, the extraction of the "financial cycle" sequence is completed.

4.2 Financial Cycle Measurement

4.2.1 Data Collection

A new Financial Cycle Index (FCI) is constructed based on the MF-DFM with text information embedding. Refer to the existing research (Goodhart & Hofmann, 2001; Rünstler, 2018), we select 37 monthly or quarterly variables of 2002 to 2022 from the three aspects: macroeconomic, monetary policy and price. Beyond these variables, 10 monthly series of processed text information are also added, which comprehensively reflect the current financial situation or sentiment. Details are shown in Table 6. M2/GDP, which is the ratio of broad money to Gross Domestic Product (GDP), usually reflects the degree of financial deepening and can be used as a reference for financial fluctuations. During periods of vigorous development, the money supply usually increases, and vice versa.

4.2.2 Data Preprocessing

A prerequisite for DFM is that the data is required to be stationary. The results of the **stationarity test** on the 47 indicators showed that the p-value are less than 0.01. We combine the Kalman filter with a dynamic factor model to deal with the missing values.

In estimating the appropriate number of factors, to ensure the results of balanced model fitting and reduce the number of parameters while ensuring explanatory power, the maximum number of factors is pre-set to 10. Results of the three information criteria are shown in Figure 8. The sequence reached its minimum value at the information criteria of ICR1, ICR2, and ICR3 at 2, 3, and 4, respectively, and then began to rise. Considering that the three criteria values and the fit of the model, the final determination of the number of dynamic factors is $q=4$.

Further, minimize the number of shocks r to the information standard, r is determined to be 3. The formula is as follows. Finally, a common factor sequence is extracted as a new FCI index to measure the financial cycle.

$$\hat{r}_n := \arg\min IC_{0;n}\left(k\right), 0 \le k \le r_{max} \tag{26}$$

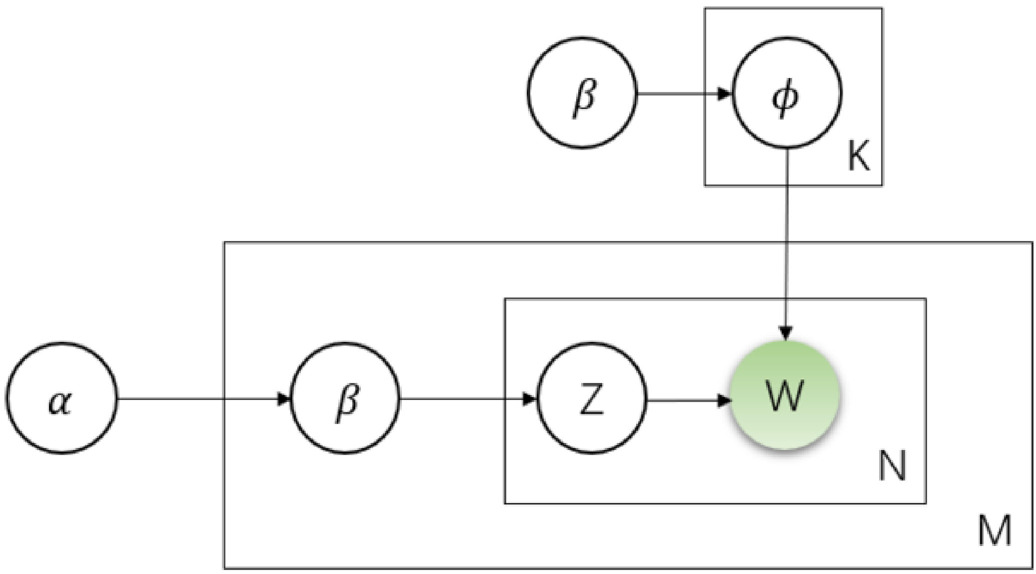
Table 4. Probability distribution of document topic after emotional adjustment

Text	Time	Stock	Policy	Regional Coop	Credit	Market Order	Globalization	Sci. & Tech	Investment	Bank	Social Finance
China will continue to implement financial policies to support the development of Tibet……	2002/1	0.0017	0.3467	0.0411	0.1509	0.0017	0.0017	0.0017	0.0017	0.0663	0.3865
Experts believe that cooperation between banks and securities is an upgrade of China's financial capabilities……	2002/1	-0.0011	-0.7264	-0.1731	-0.0680	-0.0011	-0.0256	-0.0011	-0.0011	-0.0011	-0.0011
Profits declined significantly due to Economy of Argentina's economic crisis……	2002/1	0.0029	0.0029	0.0029	0.4419	0.0029	0.3052	0.0029	0.1298	0.1055	0.0029
……	…	…	…	…	…	…	…	…	…	…	…
The financial market has not entered the "once in a decade" fluctuation cycle……	2022/12	-0.1668	-0.0008	-0.0008	-0.3083	-0.0356	-0.3970	-0.0008	-0.0008	-0.0448	-0.0445

Table 5. "Financial cycle" text information series

Time	Stock	Policy	Regional Coop	Credit	Market Order	Globalization	Sci. & Tech	Investment	Bank	Social Finance
2002-01	0.0005	0.2145	0.2428	0.0510	0.0005	-0.0445	0.0005	0.0179	0.0396	0.0774
2002-02	0.0460	0.1587	0.1814	0.0324	-0.0332	0.0327	0.0059	0.1206	0.0459	0.0226
2002-03	0.0150	0.1107	0.1819	0.0377	-0.0324	0.1085	0.0038	0.0978	0.0542	0.0111
...	...	...	...	...	...	...	...	...	...	...
2022-10	0.0364	0.1750	0.2151	0.0722	0.0034	0.0740	0.0697	0.0590	0.0269	0.1472
2022-11	0.0138	0.0842	0.0937	0.0417	-0.0094	0.1121	0.0165	0.0769	0.0172	0.0323
2022-12	0.0556	0.0812	0.0985	0.0539	0.0081	0.0551	0.0137	0.1642	0.0417	0.0575

Figure 7. Series fluctuations of 10 topics in LDA



$$IC_{0;n} := \frac{1}{n}\sum_{j=k+1}^{n}\int_{-\pi}^{\pi}\lambda_{nj}\left(\theta\right)d\theta + kr\left(n\right) \tag{27}$$

4.3 Dynamic Factor Extraction

4.3.1 The Construction of the New FCI

M2/GDP is considered as a reference for traditional financial cycle measurement (the dashed line), and some practical significant events are used to calibrate the measures. As shown in Figure 9, FCI constructed by the **MF-DFM with embedded text information** (the solid line) can well reflected in the financial crisis in 2008, the stock market crash in 2015 and the forecast for COVID-19 in 2020, by generally advance nearly one year compared to M2/GDP. Therefore, it is preliminarily believed that FCI measured using this method has a certain degree of accuracy. Meanwhile, as an indicator that incorporates a wide range of information, is more sensitive to the subtle fluctuations in the financial markets and is therefore more predictive and accurate than the M2/GDP, by more subtly presenting the peaks and valleys of the cycle for typical financial events.

To demonstrate the effectiveness and accuracy of the FCI with embedded text information, we further adopt the two strategies:

(1) Compare the (a) FCI with embedded text information with the (b) index with pure text information and (c) FCI with traditional indicators. If the (a) is found to be more fitted in the long-term trend, it is considered to be superior.
(2) Compare the changing trends of (a) FCI with embedded text information with economic indicators such as GDP growth rate and CPI. If the trends are similar, it can be considered that (a) responds well to the financial cycle. If the trend is ahead, it can be considered that (a) can predict the financial cycle.

Table 6. Building indicators for mixed frequency dynamic factor models in financial cycles

Index	Variable	Definition	Frequency
Macro-economics	shanghai_index	Shanghai Composite Index_ Closing_ Current period	monthly
	loan	Shanghai Composite Index_ Closing_ Current period	monthly
	social_finance	Incremental scale of social financing_ Cumulative	monthly
	deposit_loan	Deposit loan ratio	monthly
	reserve	Foreign exchange reserves_ End of period	monthly
	shenzhen_1	Shenzhen Composite B-share Index_ Closing	monthly
	shenzhen_2	Shenzhen Composite A-share Index_ Closing	monthly
	loan_rate_1	Financial institutions' benchmark interest rate for RMB short-term loans within 6 months	monthly
	loan_rate_2	The benchmark interest rate for RMB short-term loans by financial institutions is 6 months to 1 year	monthly
	share	Year-over-year change rate of stock price index	monthly
	gdp	gross national product	quarterly
	cpi	consumer price index (base year=100)	monthly
	stock	Year-over-year change in stock price index	monthly
	Indus_confi	Industrial confidence indicators	monthly
	business_confi	Business confidence indicators	monthly
	confidence	Consumer confidence indicators	monthly
monetary policy	exchange_rate	Real effective exchange rate index	monthly
	bank_1	Weighted average interest rate of inter-bank Interbank lending market	monthly
	bank_2~bank_5	Weighted average interest rate of inter-bank Interbank lending market: 7 days, 1 month, 2 months, 3 months	monthly
	m0	M0 (circulating cash in the market)	monthly
	m1	M1 (narrow Money supply)	monthly
	m2	M2 (Broad money supply)	monthly
	m3	M3 (Practical and Total Borrowing)	monthly
price system	housing	National Housing Prosperity Index_ Current period	monthly
	oil	Average price of crude oil in the international market	monthly
	ppi	Year-on-year change rate of industrial price index	monthly
	ppi_out	Producer Factory Price Index	monthly
	ppi_in	Industrial Producer Purchase Price Index	monthly
	inflation	Overall inflation rate (%)	quarterly
world market	export	Year-on-year growth rate of export volume	monthly
	import	Year-on-year growth rate of import volume	monthly
	ca	Current account balance (USD)	quarterly
	fa	Financial account difference (USD)	quarterly
text information	topic1-10	Stock, Policy, Regional Coop, Credit, Market Order, Globalization, Sci. & Tech, Investment, Bank, Social Finance	monthly
	m2gdp	M2/GDP - Degree of Financial Deepening	quarterly

Data sources: China Economic Network Statistical Database, National Bureau of Statistics, OECD Database

Figure 8. Results of information criterion

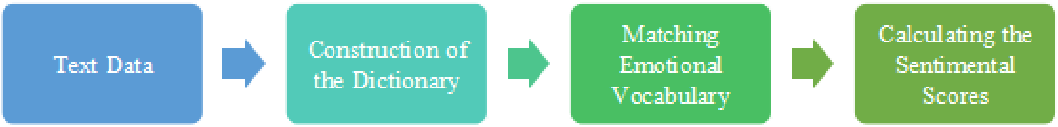
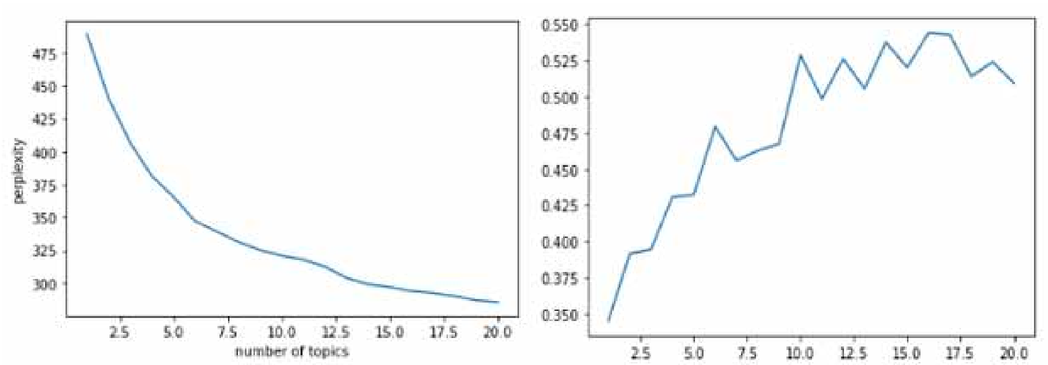


Figure 9. Fitting results of embedded text information indicators



4.3.1 Internal Comparison: Advantages of Embedding Textual Information

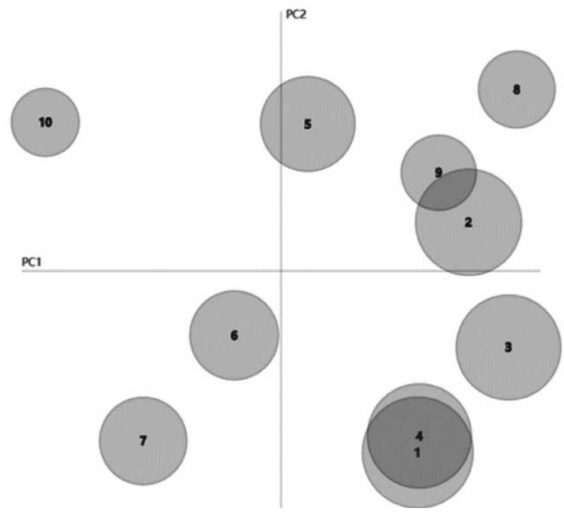
The series with 10 text indicators and the 37 traditional indicators were reconstructed separately by DFM following the steps described above. Their common factor series were taken as indices to help assess the financial cycle. The 3 series are presented in quarterly form in Figure 10. It can be observed that (b) index with pure text information can depict the financial cycle to some extent. (c) FCI with traditional indicators also reflects the financial cycle well and has significant characterizations at important time nodes (2008, 2015, 2020). Compared with (b) and (c), (a) FCI with embedded text information reflects fluctuations and changes in more detail, and it is more forward-looking at major time points such as the 2008 financial crisis and the COVID-19. The R^2 of the (a) to observation series is 0.6126, while the R^2 of (c) observation series is 0.5601, which means the traditional indicators of (c) has a certain lag. Comparison shows that only the (a) FCI with embedded text information is most consistent with the trend of the traditional economic indicators and has the strongest precedence, which suggest that financial cycle measurement that combine text information and traditional indicators is of more advantageous.

4.3.2 External Comparison-Precedence

Compare the changing trends of (a) FCI with embedded text information with economic indicators such as GDP growth rate and CPI. The changing trend of GDP growth rate is taken as the quarterly growth rate and the changing trend of CPI is the average of the monthly data transformed into the quarterly data.

The GDP growth rate can be seen as the economic growth and, in general, the higher the level of economic growth, the more the financial institutions are willing to take risks and expand lending, which in turn boosts financial markets. While CPI reflects the condition of inflation, which may lead to higher asset prices, affecting market sentiment and investor confidence, and is thus closely linked to the financial cycle. CPI varies at different stages of the financial cycle. As shown in Figure 11, to facilitate the observation, the two indices CPI and GDP are normalised. As can be seen from the graph, the indicators show a more consistent trend of fluctuation across all stages, except for

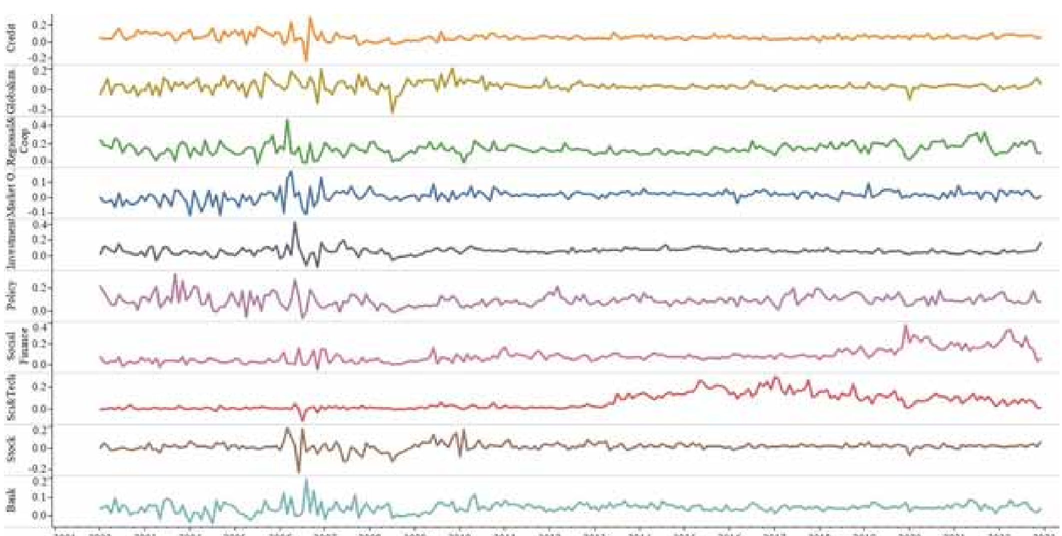
Figure 10. Comparison of mixing dynamic factor model indicators composed of different data components (monthly)



around early 2002 to 2003 and 2021, when the indicators show heterogeneity fluctuations. (a) FCI with embedded text information can better depict the 2015 stock market crash and the 2008 financial crisis than the GDP growth. Compared to the CPI growth, (a) presented a 2-4 period lead time. It also shows that (a) have a good form of convergence with the macroeconomy. Specifically, during the prosperous cycles, the macro tends to expand, while during periods of deteriorating financial conditions, it tends to decline.

Furthermore, we conducted a Granger Causality Test of the growth on (a), GDP growth rate and CPI. As is shown in Table 7, based on the results in the table, it can be seen that growth on (a) can cause growth on GDP and CPI, which also indicates that (a) is ahead of these indices and can play a certain early warning role on macro trends as well. As (a) covers information from several financial

Figure 11. Comparison of growth rates of FCI, CPI, and GDP (quarterly)



markets, including the stock, bond and foreign exchange market, it can reflect the overall condition of the financial market more comprehensively.

4.4 Nowcasting by FCI

4.4.1 Nowcasting the Financial Cycle

In addition to extract common factor series, (a) FCI with embedded text information with mixed-frequency indicators is also able to perform Nowcasting of the target series. As it is able to extract common time series changes rather than treating the time series of each indicator separately, which allows it to capture macro cycles and trends. Therefore, use (a) to forecast the financial cycle trend for the next 12 months using a mixed-frequency dynamic factor model for its immediate prediction. The model was constructed in the same steps as above and the forecast results are shown in Figure 12.

4.4.2 Analysis of the Financial Circle

Based on the results of the measurement and forecasting of (a), the financial cycle is defined and classified based on the turning point analysis. The time horizons of the short, medium and long cycles are defined for the *Kitchin, Juglar* and *Kuznets* cycles respectively, as shown in Table 7.

Due to the limited time span, analysis of medium and long cycles is difficult to carry out, and therefore the short cycles analysis is executed. Referring to periods setting in Drehmann et al. (2012),

Table 7. Glenmorangie distillery causality test between FCI and economic climate index

Paired Sample		F	P	Conclusion
FCI	CPI	3.547	0.011**	FCI can cause changes in CPI
CPI	FCI	0.654	0.626	CPI cannot cause changes in FCI
GDP	FCI	2.322	0.065*	GDP can cause changes in FCI
FCI	GDP	7.391	0.000***	FCI can cause changes in GDP

Note: * * *, * *, * represent significance levels of 1%, 5%, and 10%, respectively

Figure 12. Prediction of FCI

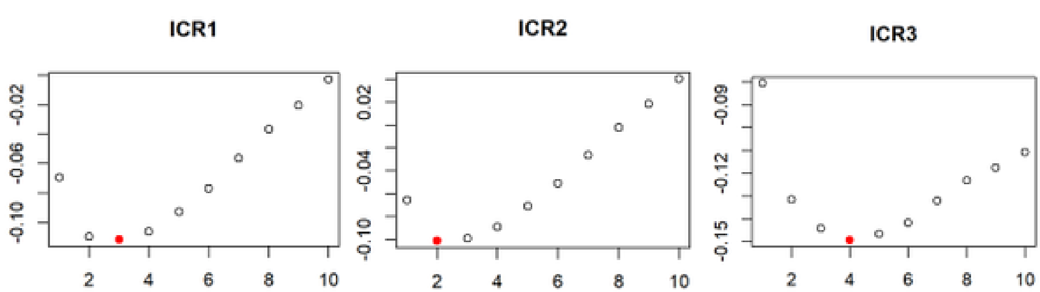


Table 8. Frequency scale division of financial cycles

Classical Periodic Theory	Cycle Length	Cycle Type
Kitchin cycle	3-5 years	short
Juglar Cycle	9-10years	medium
Kuznets Cycle	15-25 years	long

the short periods were determined to be (1) a minimum duration of 6 months for expansion and contraction periods (2) a minimum duration of 15 months for a full cycle.

Figure 13 shows the volatility of the financial cycle carved out using (a) and its forecasts. The total volatility series shows that the financial cycle can be divided into long and short cycles. Table 8 shows the corresponding characteristics.

As can be seen in Table 9, FCI measures six troughs and six peaks from January 2002 to December 2023 including forecasts, with five complete short financial cycles based on the "peak (trough)-peak (trough)" classification criteria. The average length of the expansion period is 18.2 months, the average length of the contraction period is 23.4 months and the average length of the financial cycle is 41.6 months. With the contraction period of the financial cycle being longer than the expansion period. The expansion which is the average growth from one trough to the next is calculated based on the concept of expansion, and vice versa for contraction. The expansion rate of the financial cycle is 257%, and the contraction rate is -240%, with an overall expansion rate greater than the contraction rate.

The financial cycle can be broadly divided into the following stages:

Figure 13. FCI characterized financial cycles (including predictions) short cycles

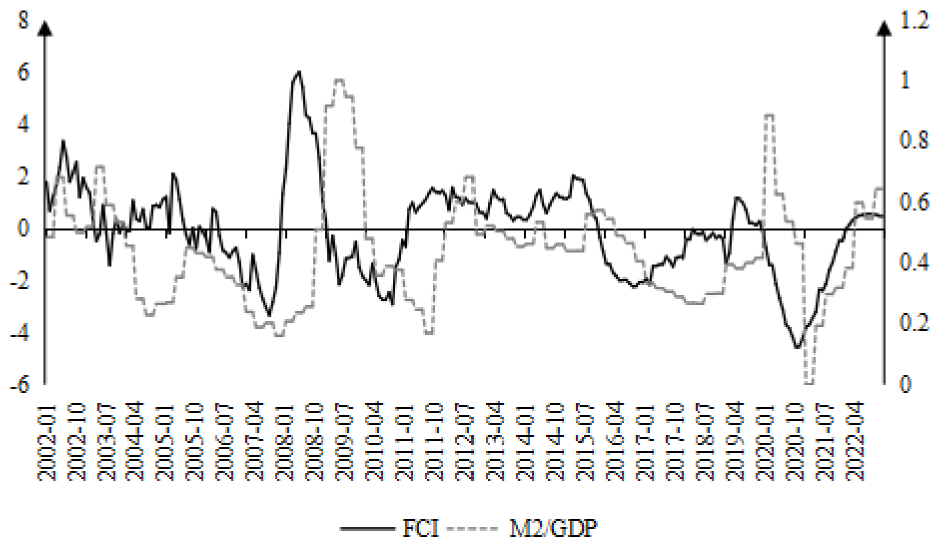


Table 9. The changing characteristics of short financial cycles

Peak	Trough	Peak	Duration			Amplitude of Vibration	
			Expansion	contraction	Length	Expansion Amplitude	Contraction Amplitude
——	2003M08	2005M03	19	——	——	257%	-240%
2005M03	2007M08	2008M05	9	29	38		
2008M05	2010M09	2011M09	12	28	40		
2011M09	2013M09	2015M03	18	24	42		
2015M03	2016M09	2019M04	31	18	49		
2019M04	2020M10	2022M07	21	18	39		
Average length			18.2	23.4	41.6		

The first stage was from 2002 to March 2005. The Central Bank implemented a prudent monetary policy and took measures to address problems such as duplication of construction, expansion of fixed investment and high prices. Inflation was controlled and economic volatility and risks were mitigated by adjusting the size of the money supply and credit. during this period, China's economy grew rapidly, the stock and real estate markets performed relatively strongly, and bank credit and money supply were relatively abundant. The FCI achieved a reversal from a reversal to an increase.

The second stage was from March 2005 to August 2007. As a result of the global financial crisis, the FCI deteriorated rapidly and fell to the lowest level in the time frame studied. At this time, the country introduced the "four trillion" stimulus policy. This led to a significant increase in the money supply and bank lending, stimulating economic recovery, which led to a rise in asset prices and a boom in the property market, and a marked improvement in the FCI.

The third stage was from May 2008 to September 2011, when, according to the data, inflation in China was higher in the run-up to the turning point and the central bank raised the reserve requirement ratio in response to inflationary risks, while higher interest rates and falling stock prices also made financial conditions tighter. During this phase, the decline in FCI was more pronounced in character.

The fourth stage was from September 2011 to March 2015, saw a period of stable financial development in the period 2013-2014 as inflation was effectively controlled and financial conditions eased.

The fifth stage was from March 2015 to April 2019, when the economy faces challenges such as the 2015 stock market crash. China experienced a series of financial policy tightening during this period, the most representative of which was the round of tightening initiated in early 2017, which was driven by quantitative variables, in particular the deleveraging of the shadow banking sector. As China's economy entered a phase of high-quality development, economic growth began to slow and uncertainty in financial policy increased significantly, and this uncertainty led to a tightening phase in the country's financial cycle. 2018 saw the beginning of a rapid easing of policy in response to concerns about a hard landing in China and the instability of the RMB exchange rate. The People's Bank of China lowered policy interest rates, deregulated finance and the shadow banking sector grew rapidly. During this period, FCI showed a decline and then an increase.

The final stage, which is the epidemic era, in the second half of 2019, the financial medium term has shown a certain downward trend, but with the spread of the new crown pneumonia epidemic in late 2019, domestic economic activity almost came to a halt and dealt a huge blow to our financial markets. The epidemic began to spread globally in February-March 2020, exacerbating the tense form of the financial cycle. This situation continued until the end of 2020, when financial conditions gradually improved as the Chinese government implemented various policies in response to the epidemic and gradually resumed work and production.

In the forecast period of 2023, from the forecast level, financial cycle will turn from up to down after 2023, but according to the forward-looking FCI model forecast, the turning point is expected to lag 2-3 quarters, i.e. it may enter the downward phase in the second half of 2023. The financial cycle will enter a certain adjustment and recovery phase, which is still unknown until the next round of expansion.

5. CONCLUSION

Re-examining the crises that occurred worldwide, financial cycle has played a key role, which has prompted research on reconsidering the interdependence between financial factors and the real economy, and thus redesign the framework of macroeconomic research, which integrate financial system as a fundamental part. To accurately grasp the financial cycle and thus achieve better economic and financial regulation, we extend the traditional financial cycle measurement by constructing a MF-DFM to achieve a new Financial Cycle Index (FCI) with financial newspaper text information embedding by using LDA. The specific results are as follows:

(1) Using information from newspaper text information from 2002 to 2022, we extracted 10 series related to the financial cycle through the LDA, covering equities, policy, regional cooperation, credit, market order, globalisation, technology, investment, banking and social finance respectively. These series not only help to better predict the trends of the financial cycle, but also to better understand the interactions and impacts between different areas in the financial cycle, providing a more accurate basis for financial risk management and policy formulation.
(2) FCI is relatively accurate. Constructed through the MF-DFM with embedded text information, FCI is an effective indicator of financial cycles, which can accurately reflect the impact of key events on financial markets, such as the financial crisis in 2008, the stock market crash in 2015 and COVID-19, which suggests that FCI contains some unique advantages in both research and application.
(3) FCI has relative superiority. By using the mixed frequency DFM with embedded text information, the index has advantages in analysis and prediction. Compared to traditional DFM, FCI can achieve a more accurate and timely characterization by considering text information on the financial cycle. Furthermore, FCI can present as an "early warning" for the financial cycle by incorporating the mixed frequency measurement, which can be found that the FCI is mostly in the lead of CPI by 2 to 4 periods, and thus indicating to be an applicable instrument for early warning.
    (5) China has experienced roughly five financial short cycles from 2002-2022, with the contraction period of the financial cycle being longer than the expansion period and the expansion being greater than the contraction. The financial cycle is characterized by different fluctuations under the influence of a complex blend of financial crisis, national policies and interest rates. Based on the immediate forecast results, the second half of 2023 is forecast to potentially see a turning point in the financial cycle.

As a better reflection on financial fluctuations, FCI can help to formulate related polices based on timely monitoring and forecasting the cycle. Text information can reflect conditions from perspectives including changes in market sentiment and hot spots, which immediately makes the measurement more sensitive, timely, and accurately. The MF-DFM can combine indicators with different frequencies as daily, monthly, quarterly frequency, thus make full use of various sources to improve the information quality and can additionally further achieve timeliness. FCI can be used as an early warning for financial cycles. By monitoring the fluctuations of FCI and the correlation with other economic indicators, it help to identify risks and take timely measures to make intervention.

FUNDING STATEMENT

This research was funded by the National Social Science Fund (18CTJ004).

REFERENCES

Aikman, D., Haldane, A., & Nelson, B. D. (2015). Curbing the Credit Cycle. *Economic Journal (London)*, *125*(585), 1072–1109. doi:10.1111/ecoj.12113

Aruoba, S. B., Diebold, F., & Scotti, C. (2009). Real-time Measurement of Business Conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427. doi:10.1198/jbes.2009.07205

Banbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, *29*(1), 133–160. doi:10.1002/jae.2306

Baruník, J., & Kehlík, T. (2017). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics*, *16*(2), 2.

Benchimol, J., Kazinnik, S., & Saadon, Y. (2022). Text mining methodologies with R: An application to central bank texts. *Machine Learning with Applications*, *8*, 100286. doi:10.1016/j.mlwa.2022.100286

Boivin, J., Kiley, M. T., & Mishkin, F. S. (2010). How has the monetary transmission mechanism evolved over time? In Handbook of monetary economics. Elsevier.

Borio, C. (2014). The Financial Cycle and Macroeconomics: What Have We Learnt? *Journal of Banking & Finance*, *45*, 182–198. doi:10.1016/j.jbankfin.2013.07.031

Bruno, V., & Shin, H. S. (2013). *Capital flows, cross-border banking and global liquidity*. National Bureau of Economic Research. doi:10.3386/w19038

Bry, G., & Boschan, C. (1972). Cyclical analysis of time series: Selected procedures and computer programs. *Journal of the Royal Statistical Society. Series A (General)*, *135*(2), 284. doi:10.2307/2344336

Claessens, S., Kose, M. A., & Terrones, M. E. (2011). How do business and financial cycles interact. *Journal of International Economics*, *87*(1), 178–190. doi:10.1016/j.jinteco.2011.11.008

Deng, C., Zhao, X., & Xu, M. (2022). Financial Cycle and the Effect of Monetary Policy. *Finance Research Letters*, *47*(6), 102570. doi:10.1016/j.frl.2021.102570

Drehmann, M., Borio, C., & Tsatsaronis, K. (2012). Characterising The Financial Cycle: Don't Lose Sight of the Medium Term! *BIS*. *Working Papers*, *68*(3), 1–18.

Feldman, R., & Sanger, J. (2007). *The text mining handbook*. Cambridge University Press.

Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series. In B. D. J. Aigner & A. S. Goldberger (Eds.), *Latent Variables in Socio-Economic Models*. North-Holland.

Ghysels, E., Rubia, A., & Valkanov, A. (2008). *Multi-Period forecasts of variance: Direct, Iterated, and Mixed-Data approaches*. Working paper, Alicante, UCSD and UNC.

Goodhart, C., & Hofmann, B. (2001). Asset Prices, Financial Conditions, and the Transmission of Monetary Policy. *Proceedings*, *114*(2), 198–230.

Hansson, B., & Lindberg, H. (1994). Monetary Conditions Index: A Monetary Policy Indicator. *The Quarterly Review*, (3), 12–17.

Harding, D., & Pagan, A. (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics*, *49*(2), 365–381. doi:10.1016/S0304-3932(01)00108-8

Hatzius, J., Hooper, P., Mishkin, F. S., Schoenholtz, K., & Watson, M. (2010). Financial Conditions Indexes: A Fresh Look after the Financial Crisis. *Working Paper Series (National Bureau of Economic Research)*, 16150. doi:10.3386/w16150

Holz, M. (2005). *A Financial Conditions Index as indicator for monetary policy in times of low, stable inflation and high financial market volatility*. Paper prepared for presentation in the 9th Workshop "Macroeconomics and Macroeconomic Policies - Alternatives to the Orthodoxy", Berlin, Germany.

Krupkina, A., & Ponomarenko, A. (2015). Balance sheet structure indicators and the financial cycle. Irving Fisher Committee on Central Bank Statistics. Indicators to support monetary and financial stability analysis: data sources and statistical methodologies. *IFC Bulletin, 39*.

Lack, C. P. (2003). A financial conditions index for Switzerland. *BIS Papers Chapters, 19*.

Mariano, B. S., & Murasawa, Y. (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics*, *18*(4), 18427. doi:10.1002/jae.695

Mimir, Y. (2016). Financial Intermediaries. *Credit Shocks and Business Bulletin of Economics and Slatistics*, *78*(1), 42–74.

Pontines, V. (2017). The Financial Cycles in Four East Asian Economies. *Economic Modelling*, *65*(C), 51–66. doi:10.1016/j.econmod.2017.05.005

Rünstler, G., Balfoussia, H., & Burlon, L. (2018). *Real and Financial Cycles in EU Countries-Stylised Facts and Modelling Implications.* Frankfurt: European Central Bank Working Papers.

Ryohei, H., Didier, S., & Takayuki, M. (2013). High Quality Topic Extraction from Business News Explains Abnormal Financial Market Volatility. *PLoS One*, *8*(6), e64846. doi:10.1371/journal.pone.0064846 PMID:23762258

Schularick, M., & Taylor, A. M. (2012). Credit booms gone busts: Monetary policy, leverage cycles, and financial crises, 1870–2008. *The American Economic Review*, *102*(2), 1029–1061. doi:10.1257/aer.102.2.1029

Silva, G F D. (2002). The impact of financial system development on business cycles volatility: Cross-country evidence. *Journal of Macroeconomics*, *24*(2), 233–253. doi:10.1016/S0164-0704(02)00021-6

Stock, J. H., & Watson, M. W. (1991). A Probability Model of the Coincident Economic Indicators. In K. Lahiri & G. H. Moore (Eds.), *Leading Economic Indicators: New Approaches and Forecasting Records* (pp. 63–89). Cambridge University Press. doi:10.1017/CBO9781139173735.005

Strohsal, T., Proaño, C., & Wolters, J. (2015). *Characterizing the Financial Cycle: Evidence from A Frequency Domain Analysis*. SFB Discussion Papers.

Wang, J., Shi, J., Han, D., & Zhao, X. (2019). Internet Financial News and Prediction for Stock Market: An Empirical Analysis of Tourism Plate Based on LDA and SVM. *Journal of Advances in Information Technology*, *10*(3), 95–99. doi:10.12720/jait.10.3.95-99

Weng, J., Lim, E. P., & Jiang, J. (2010). Twitter Rank: Finding topic sensitive influential twitterers. *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, 456-463.