2025 Volume 3, Issue 10
Lecture Notes in Education, Arts, Management and Social Science
ISSN(Online): 2705-053X

Real-time Forecasting of Business Climate Index in Cultural, Sports, and Entertainment Industries of China: A Mixed-Frequency Dynamic Factor Model Approach

Lina Huang, Qiwen Chen*, Houzhong Jin
School of Sport Economics and Management, Central University of Finance and Economics, Beijing 102200, China

**Author to whom correspondence should be addressed.*

Abstract: Since China's economic reform, the cultural, sports, and entertainment (CSE) industries have experienced significant growth. Currently, there is a lack of effective high-frequency indicators to help policymakers and industry practitioners monitor CSE developments in real time. This study constructs a Mixed-Frequency Dynamic Factor Model to provide real-time forecasting of the Prosperity Index of Enterprises (PIE) in China's CSE industries, utilizing a dataset consisting of 26 macroeconomic indicators from 2010 to 2024. The results revealed that the model effectively captured fluctuations in PIE, successfully distinguishing economic situation before and after the COVID-19. Compared to existing macroeconomic forecasting models, this model exhibits superior predictive accuracy.

Keywords: Dynamic factor model; Cultural industry; Sports industry; Entertainment industry; China; Mixed-frequency data; Real-time forecasting

Online publication: November 26, 2025

1. Introduction

Since the initiation of the Reform and Opening-up policy, China's economy has experienced rapid growth. The country's share of the global economy has steadily increased, and in 2010, China officially surpassed Japan to become the world's second-largest economy. Market-oriented reforms have injected new vitality into the Chinese economy—China's stock market has evolved from nonexistence to one of the most significant capital markets worldwide; foreign trade has achieved remarkable progress, with China maintaining its position as the world's largest goods trading nation for eight consecutive years; and industrial restructuring has continued to advance, with the tertiary sector experiencing robust growth. Notably, industries such as culture, sports, and entertainment have expanded significantly, emerging as new drivers of economic growth.

The rapid rise of China's economy has significantly increased the disposable income of urban and rural residents, leading to a substantial rise in both their acceptance of and participation in discretionary consumption [1]. Driven by a massive consumer base and supportive national policies, China's tertiary sector has demonstrated immense growth potential, with its scale expanding annually. Since 2023, following the full reopening in the post-pandemic era, China's cultural, entertainment, and sports industries have experienced a remarkable recovery and growth. In the cultural and entertainment sector, the concert market has rebounded rapidly, with China's total concert box office revenue surging in 2024. The gaming industry has also seen positive developments, with China's first domestically produced 3A game, *Black Myth: Wukong*, receiving unprecedented attention and widespread acclaim. In the sports industry, major events such as the 2022 Winter Olympics have fueled the rapid growth of China's winter sports industry. As one of the most essential components of discretionary consumption, the cultural, sports, and entertainment industries have fully benefited from the market-driven economy, playing a pivotal role in economic transformation and consumer demand stimulation.

1.1. The cultural, sports, and entertainment industries in China: History, concepts, and development

Before the 1970s, China operated under a planned economy, where public entertainment and sports activities were primarily provided as government services. The cultural, sports, and entertainment industries relied on public funding, and professions such as athletes, actors, and singers were classified as state employees receiving fixed salaries from the government. Following the Reform and Opening-up policy, these industries gradually transitioned away from their public service nature and moved toward marketization [2]. As China's economy became increasingly market-driven, these industries embraced commercialization, expanding in response to rising consumer spending. According to the National Bureau of Statistics of China (NBS), the GDP of the cultural, sports, and entertainment industries reached 857.82 billion yuan in 2022, reflecting a year-on-year growth of 0.97%

In 2002, NBS issued the Industrial classification for national economic activities (GB4754-84), defining the Cultural, Sports, and Entertainment (CSE) industry as 'a comprehensive sector aimed at meeting people's spiritual and cultural needs while promoting the coordinated development of the economy and culture. This industry encompasses cultural creation and dissemination, sports events and fitness services, as well as entertainment and leisure activities, with culture, sports, and entertainment as its core components' [3]. In the latest classification of GB/T4754, CSE was formally classified as an independent category, consisting of five major sectors: news and publishing, broadcasting, television, film, and audio production, cultural arts, sports, and entertainment.

With increasing domestic emphasis on CSE industry, the statistical data regarding CES has been continuously refined, with new categories and indicators gradually being incorporated. However, the public disclosure of CSE industry data in China remains relatively limited. Despite these constraints, both domestic and international scholars have conducted research on CSE industry based on available data, providing valuable insights for investors, industry professionals, and policymakers. Jin analyzed the total factor productivity contribution and convergence patterns of CSE industry using data from 31 provinces (municipalities and autonomous regions) in China from 2005 to 2017 [4]. He et al. measured the total factor productivity of the cultural industry from 2000 to 2016 [5]. More recently, Liu conducted a quantitative analysis of China's sports industry. Some scholars have also approached the study of China's CSE industry from a forecasting perspective, attempting to capture its development trends [6]. Yang examined the growth patterns of China's sports industry using big data analytics and proposed a sports industry growth prediction model based on a genetic neural network [7]. Similarly, Hu focused on consumption and revenue trends in China's sports industry, proposing forecasting frameworks based on these observed patterns [8].

Currently, most research on the development of China's CSE industry is based on annual data. From a data perspective, the available data is quite limited and lacks timeliness in updates. While China's GDP growth rate is released quarterly, the quarterly contribution of CSE industry is not disclosed, which hampers the ability to provide timely decision-making information for production and investment activities. When businesses plan production scales or adjust their operations, they lack precise quarterly data guidance, which increases the risk of decision-making errors. Similarly, investors face challenges in making informed investment decisions due to the absence of critical data, making it difficult to accurately identify investment opportunities and directions. These issues have created certain obstacles for the development of CSE industry, as well as its related sectors.

1.2. Prosperity Index on Enterprises

The Prosperity Index on Enterprises (PIE) is defined as an index compiled through the Enterprise Prosperity Survey (EPS), which collect business operators' subjective judgments and expectations regarding the macroeconomic environment and the operational conditions of their enterprises. This index is published quarterly and quantifies business performance, providing a dynamic reference for economic analysis. It reflects businesses' perceptions and confidence in the macroeconomic environment. Since the third quarter of 2019, China has been publishing the PIE for CSE industry and its sub-sectors. Among the existing statistical data related to CSE industry, this indicator is a reliable indicator that reflects the overall development of the industry.

The EPS originated in Western countries during the 1920s and has since been rapidly adopted and widely implemented around the world. In China, the EPS is conducted on a quarterly basis and primarily consists of qualitative questions. The questionnaire primarily includes the following common questions: judgment and expectations regarding the enterprise's operational conditions, assessment and outlook of the overall performance of the enterprise's industry, expectations concerning the domestic macroeconomic situation, judgment of dynamic changes in indicators such as enterprise profits, labor plans, investment plans, and orders. Industry-specific questions are tailored to reflect the unique characteristics of each sector.

The PIE is derived from the results of the EPS. Generally, the index is a composite calculated by weighting the current PIE (reflecting the present economic situation) and the expected PIE (reflecting future economic trends). The PIE ranges from 0 to 200, with 100 serving as the threshold. A value above 100 indicates that the economy is in an expansionary phase, with a higher number indicating greater optimism, whereas a value below 100 signals economic contraction or the spread of pessimism. Since the PIE is based on businesses' feedback regarding key indicators such as the current operational environment, orders, and profitability, compared with the same period from the previous year, it carries a certain year-on-year characteristic. This allows it to reflect the differences between the current economic situation and the same period in the previous year, helping to analyze the changing trends in economic performance.

As an important indicator for assessing the health of industries and enterprises, the PIE has been widely applied and studied in both domestic and international literature. For example, PIE can serve as a leading indicator for macroeconomic trends. Han through analyzing quarterly data from 2001 to 2011, found that PIE is highly correlated with the year-on-year GDP growth rate of the next quarter, meaning that PIE can predict economic changes about one quarter in advance and effectively capture economic turning points [9]. The *ifo* Business Climate Index in Germany is similar to China's PIE. Lehmann confirmed that the *ifo* index has high predictive power for Germany's GDP, particularly with respect to the three main components of the *ifo* index: business climate, business situation, and business expectations [10]. Additionally, PIE can be used to quantify the impact of policy shocks. For instance, Yang in the context of tax administration reforms during China's Fourteenth Five-Year Plan, constructed the key tax source PIE to predict changes in tax revenues and trends [11]. Various studies have attempted to innovate and optimize PIE prediction models. For example, Shi et al. used GMDH and RGMDH neural network methods to predict China's PIE and entrepreneur confidence index [12].

Although significant progress has been made in research related to PIE, the models used to predict it are still relatively limited, and often fail to meet the increasingly complex analytical demands. To address this gap and further improve the accuracy and adaptability of predictions, this study innovatively introduces the dynamic factor model. By using a macroeconomic dataset, we aim to conduct a more in-depth forecasting and analysis of PIE.

1.3. Dynamic factor model

Unlike traditional time series forecasting, which typically requires balanced panel data, the economic dataset used in this study has several unique characteristics. First, PIE of CSE industry is updated quarterly, while most of the economic indicators related to CSE industry are reported on a monthly basis. As a result, the dataset used for forecasting includes a mix of monthly and quarterly variables, referred to as mixed-frequency data. Second, due to constraints in data collection and measurement methods, NBS data releases have a certain lag period, which means that PIE statistics for the previous quarter cannot be published immediately at the beginning of the next quarter. In fact, regardless of whether the data is monthly or quarterly, there is typically a lag of 1 to 3 weeks in its release. This causes traditional economic data that is updated in real-time to exhibit ragged-edge characteristics at the start of each month, which significantly impacts the temporal integrity of the data. Additionally, due to factors such as the Chinese New Year holiday and specific statistical rules, monthly statistical indicators in China generally experience missing data in January.

DFM, as one of the mainstream mixed-frequency modeling methodologies, exhibit unique methodological characteristics and application paradigms in macroeconomic real-time forecasting. It has been extensively adopted in authoritative prediction systems such as the Federal Reserve Bank of Atlanta's *GDPNow* and the New York Fed's *Nowcasting*, establishing themselves as critical technical tools for forecasting core macroeconomic indicators like GDP growth rates [13,14]. Grounded in the latent factor-driven hypothesis, the model achieves data dimensionality reduction and systemic dynamic feature extraction by mapping high-dimensional observable economic variables into a low-dimensional latent factor space. Its theoretical framework comprises two core modules: the observation equation establishes a common factor extraction mechanism through macroeconomic variables, while the factor state transition equation, based on time series modeling assumptions, employs Kalman filtering and smoothing algorithms to dynamically track and update latent factor states.

Current domestic research in China employing MF-DFM for real-time forecasting remains relatively limited. Wang et al. pioneered the first single-factor MF-DFM framework incorporating six macroeconomic variables, achieving continuous estimation and forecasting of China's monthly year-on-year GDP growth rates [15]. Zheng et al. further advanced this by overcoming dimensionality constraints, validating the model's predictive validity for GDP growth during economic stabilization and post-pandemic recovery periods through a dual-factor mean squared error-based variable screening mechanism applied to a high-dimensional dataset comprising 411 economic indicators [16]. Meanwhile, Yi et al. implemented critical algorithmic enhancements by embedding Lasso variable selection techniques into the EM parameter estimation process. This innovation not only improved GDP forecasting accuracy but also established a macroeconomic risk monitoring system based on information shock decomposition [17].

In addition to forecasting GDP growth rates, DFM and its derivative models have also been applied to predict other macroeconomic indicators. Kabundi et al. used a combination of principal component analysis and Kalman filtering in DFM to forecast South Africa's financial conditions index [18]. Algaba et al. employed MF-DFM for real-time forecasting of Belgium's consumer confidence index [19]. Zheng et al. used daily-frequency MF-DFM to predict the economic uncertainty index [16]. These models have all achieved good fit results, which, to some extent, demonstrate the applicability of MF-DFM to macroeconomic indicators other than GDP.

1.4. Aims of the study

The aims of our study are to use the DFM to conduct real-time forecasting and model analysis of PIE for China's CSE industry. First, relevant macroeconomic indicators for the CSE industry are selected to fit the MF-DFM, and the forecast values generated by the MF-DFM are used to estimate the actual PIE. Second, based on the smoothed estimates produced by the model, the ability of the MF-DFM to capture past information of PIE is observed. Finally, the out-of-sample forecast results of the model are compared with those of traditional time series econometric models to assess the forecasting capability of the model.

2. Method

2.1. Model construction

2.1.1. Dynamic factor model

The dynamic factor model (DFM) underpinning this study posits that macroeconomic dynamics are driven by r latent factors, which are assumed to exhibit time-varying characteristics. To capture this temporal variation, the time-evolving dynamics of the common factors are modeled using a first-order vector autoregressive process:

$$f_t=Af_{t\text{-}1}+\epsilon_t,\epsilon_t\sim\text{i.i.d. }N(0,Q)$$

Where represents f_t-dimensional vector of latent common factors, $A_{r\times r}$ denotes the autoregressive coefficient matrix, and ϵ_t is the state transition noise with covariance matrix $Q_{r\times r}$, which characterizes the endogenous uncertainty of the latent factor system. The model assumes that the state noise follows a normal distribution and is independent of the observation noise, forming a complete Gaussian linear state-space system.

For the unobservable latent factors, their estimates can be refined using observed data. Specifically, observed variables are utilized to extract factors and update latent estimates at each time y_t (t=1,…,T). Let (denote the observed variables, whose dynamic features are dimensionally reduced through the latent factor space via the observation equation:

$$y_t=Cf_t+\mu_t,\mu_t\sim\text{i.i.d. }N(0,R)$$

Where $C_{n\times r}$ is the factor loading matrix, capturing the linear relationships between observed variables and latent factors. The term $C\times f_t$ represents the portion of y_t explained by common factors. The idiosyncratic noise μ_t, with covariance matrix $R_{n\times n}$, is constrained to be diagonal, implying that the noise components across variables are independently and identically distributed normal with no cross-sectional correlations.

By combining the observation equation in equation (2) with the state transition equation in equation (1), a state-space model can be constructed. Given the information set $\Omega_t=\{y_1,y_2,\ldots,y_t\}$ composed of the observation data, the Kalman filter and smoother can be used to estimate the latent factors at any given time.

To address the prevalent issue of missing values in macroeconomic data, this paper employs a structured processing scheme based on the state-space model. By introducing a missing indicator diagonal matrix W_t. A data integrity discrimination mechanism is established: when the i-th component of the observed variable. y_t contains a valid value, the corresponding diagonal element of W_t is set to 1; when the component is missing, the diagonal element is set to 0. By left-multiplying both sides of Equation (2) with W_t. The interference of missing data on state estimation is systematically eliminated, forming a standardized framework for missing data processing.

In terms of parameter estimation, this study employs the iterative optimization strategy of the EM algorithm, with its technical approach adhering to the quasi-maximum likelihood estimation framework proposed by Doz et al. [20] The parameter set to be estimated, A,C,Q,R,Z_0,P_0, includes the structural parameters of the state-space model and the initial conditions for Kalman filtering. The algorithm decomposes the joint probability density function and constructs the conditional expectation based on the complete-data log-likelihood:

$$\mathrm{L}(\theta)=\mathbf{E}\quad[\log p(y_{1:T}|f_{1:T},C,R)]+\mathbf{E}\quad[\log p(f_{1:T}|A,Q,f_0,P_0)]$$

By maximizing the expected log-likelihood function, the parameter update formulas can be derived.

2.1.2. Mixed-frequency data processing

When processing mixed-frequency data through a mixed-frequency dynamic factor model, quarterly data are treated as a monthly series with observations existing only in the third month of each quarter. By establishing a conversion relationship between the latent monthly series and the quarterly series, the model integrates data of different frequencies by introducing lagged factor terms into Equation (2). For quarter-over-quarter growth data, Mariano et al. [21] proposed a correspondence formula between quarterly series and 4th-order lagged monthly series, which was applied by Bok et al. to the New York Fed’s *Nowcasting* forecasting model [13]:

$$y_t=\frac{1}{3}y_t^*+\frac{2}{3}y_{t\text{-}1}^*+y_{t\text{-}2}^*+\frac{2}{3}y_{t\text{-}3}^*+\frac{1}{3}y_{t\text{-}4}^*$$

Where Y_t^* represents the latent month-on-month observation series. For year-on-year data, Zheng et al. [22] demonstrated the approximate relationship between the quarterly year-on-year series and the corresponding monthly year-on-year series as :

$$y_t = \frac{1}{3} y_t^* + \frac{1}{3} y_{t-1}^* + \frac{1}{3} y_{t-2}^*$$

Applying the above conversion relationship to our model, using a bivariate model consisting of a single monthly indicator y_t^M and a single quarterly indicator y_t^Q as an example, the state-space equation can be rewritten as:

$$\begin{pmatrix} y_t^M \\ y_t^Q \end{pmatrix} = C \begin{pmatrix} f_t \\ \frac{1}{3}(f_t + f_{t-1} + f_{t-2}) \end{pmatrix} + \mu_t$$

At this point, the quarterly indicator is treated as a latent monthly indicator, with the left-hand side of the equation unified to a monthly frequency. By setting the corresponding structure for the state transition matrix , a mixed-frequency model incorporating both monthly and quarterly indicators can be established.

2.1.3. Model structure

Building on the methodological framework established by Bańbura et al. [23], the model incorporates idiosyncratic components to capture residual fluctuations in macroeconomic indicators that remain unexplained by common factors. At the technical implementation level, the temporal correlation of idiosyncratic components is modeled via processes, with the default assumption of no cross-correlation among these components across indicators. However, given that the dataset involves CSE interconnected indicators, endogenous correlations may arise in idiosyncratic components due to industrial linkages. To address this, the study references the factor-blocking method proposed by Bok et al. designing an additional factor block specifically for CSE industry-related data to effectively control cross-indicator correlations. Furthermore, adhering to the parameter constraints outlined by Bańbura and Modugno the model assumes independence among observation errors with fixed variances set to 10^{-4}, balancing model complexity and numerical stability in parameter estimation.

In the actual modeling process, it is assumed that within each factor block, the macroeconomic variables are driven by a single latent factor, with the factor satisfying first-order serial correlation. Therefore, within each factor block, the number of factors is set to 3, which includes one latent factor satisfying first-order lag and the second-order lag of this latent factor (to satisfy the quarterly frequency conversion requirement). After parameter initialization, the Kalman filter and smoother are used for state estimation, followed by parameter optimization using the EM algorithm. This process is iterated repeatedly until the convergence condition is met (i.e., the change in the log-likelihood value is less than 10^{-4}).

Let PIE be the last indicator of y_t, denoted as $y_{n,t}$. After optimizing with the EM algorithm, the smoothed value of PIE in the complete time series, $\hat{y}_{n,t}$, can be obtained using the factor values from the last Kalman smoothing and the loading matrix .

$$\hat{y}_{n,t} = Cf_{t|T}, \quad t = 1:T$$

To make a forecast for a given quarter, it is sufficient to estimate the smoothed forecast values for the three latent monthly values of that quarter in the form of a state-space model. The forecast can then be obtained by computing the weighted average of the three forecast values according to equation (5).

2.2. Data description

The study selects a total of 26 macroeconomic indicators across various industries. Among them, the cumulative completed floor area of construction enterprises of CSE industry, the cumulative completed value of construction enterprises of CSE industry, and the PIE of CSE industry are quarterly data, while the remaining industry economic indicators are presented monthly. In order to take into account the availability of data and the diversity of data involved, this paper selects price-type, industry boom-type, foreign trade-type, industry domestic demand-type, enterprise development-type indicators, as well as industry stock market indicators, and strives to comprehensively reflect the development of China's CSE industry. The data sample in this paper spans from January 2010 to December 2024, with the main data sourced from the NBS and other indicators refined from CEInet statistics Database. In accordance with model specification requirements and following the conventional practice of Chinese statistical data release, the year-on-year growth rates of the respective economic indicators are employed as the research variables. The selected industry economic indicators are summarized in **Table 1**.

Table 1. Macroeconomic indicators

Index	Frequency	Culture	Sport	Entertainment	Source
Industrial Producer Price Index	m	0	0	0	NBS
Stock Transaction Amount of CSE(SSE)	m	1	1	1	NBS
Stock Transaction Amount of CSE(SZSE)	m	1	1	1	NBS
Fixed Asset Investment of CSE	m	1	1	1	NBS
Export Delivery Value of Cultural Educational Sports and Entertainment Industry	m	1	1	1	NBS
Liabilities of Cultural Educational Sports and Entertainment Enterprises	m	1	1	1	NBS
Profits of Cultural Educational Sports and Entertainment Enterprises	m	1	1	1	NBS
Operating Costs of Cultural Educational Sports and Entertainment Enterprises	m	1	1	1	NBS
Operating Revenue of Cultural Educational Sports and Entertainment Industry	m	1	1	1	NBS
Assets of Cultural Educational Sports and Entertainment Industry	m	1	1	1	NBS
Industrial Value Added of Cultural Educational Sports and Entertainment	m	1	1	1	NBS
Weighted Average Turnover Rate of CSE(SSE)	m	1	1	1	SSE
Completion Value of CSE Buildings	q	1	1	1	NBS
Completion Area of CSE Buildings	q	1	1	1	NBS
Consumer Price Index of CSE (Current)	m	1	0	1	NBS
Private Fixed Asset Investment in CSE	m	1	1	1	NBS
Export Amount of Sports Goods and Equipment	m	0	1	0	GAC
Export Amount of Cultural Products	m	1	0	0	GAC
Import Amount of Cultural Products	m	1	0	0	GAC
Total Retail Sales of Sports and Entertainment Goods by Enterprises Above the Limit	m	0	1	1	NBS
Total Retail Sales of Cultural Goods by Enterprises Above the Limit	m	1	0	0	NBS
Consumer Employment Expectation Index	m	0	0	0	NBS
Consumer Income Expectation Index	m	0	0	0	NBS
Consumer Confidence Index	m	0	0	0	NBS
General Public Budget Expenditure on CSE	m	1	1	1	MOF
PIE	q	1	1	1	NBS

Note: Culture, sports and entertainment respectively indicate that in the MF-DFM estimation process, the variables selected by the industry (variable = 1) are loaded into the corresponding factor block to carry out correlation control within the industry. m: monthly series; q: quarterly series; NBS: National Bureau of Statistics of China; SSE: Shanghai Stock Exchange; GAC: General Administration of Customs of China; MOF: Ministry of Finance of China.

In real-world scenarios, the release dates of economic indicators are not uniform. Consequently, we obtained the complete dataset in early February 2025, which included the official statistics for all indicators up to December 2024 (or 2024:Q4). Based on this complete dataset, we employed a rolling cut-off strategy using monthly intervals to generate 24 backward rolling datasets, extending back to December 2021. This approach was designed to simulate the real-time nowcasting process for out-of-sample predictions. This procedure effectively assumes that in an actual real-time nowcasting scenario, when predicting the following month's data, all statistical data for the previous month has already been acquired. Additionally, since quarterly variables are assigned to the final month of the corresponding quarter within the model, the prediction process assumes that current-quarter data is available when making predictions in the final month of each quarter. Thus, the forecasting target corresponds to the data for the subsequent quarter.

Among all the time series variables, the PIE of CSE has the latest statistical starting point, beginning in the third quarter of 2019. The operating costs and revenue of enterprises in the education, culture, sports, and entertainment sectors have been recorded since February 2017, while the Consumer Employment and Income Expectation Index has been available since December 2016. To ensure the adequacy of time series data within the model, this study initiates the out-of-sample real-time nowcasting simulation for the year-on-year growth rate of PIE from the first quarter of 2022, utilizing statistical data up to December 2021. As time evolves and new data are released, the EM algorithm re-estimates the MF-DFM based on the newly acquired data and updates the out-of-sample forecasts for the PIE. Consequently, at multiple time points within a given quarter, nowcasts for the current quarter's PIE (in the first and second months) or the next quarter's PIE (in the third month) can be obtained.

3. Results

3.1. Model fitting results

First, the effectiveness of the MF-DFM in capturing the trend of macroeconomic indicators was examined. The MF-DFM was constructed using the complete dataset spanning from January 2010 to December 2024 to observe the smoothed trend of PIE across the entire time series. To obtain the quarterly smoothed estimate series, the latent monthly series of PIE was weighted and averaged within each quarter according to the conversion weights from the latent monthly variable to the quarterly variable. This estimated series was then compared with the actual PIE values published by the NBS to assess whether the model can reasonably capture past information (**Figure 1**). The results showed that the two exhibited a high degree of similarity, and the model was able to effectively capture the trend of the PIE, even during special periods such as the COVID-19 outbreak (Q1:2020–Q2:2020) and the post-reopening period (Q1:2023–Q3:2023), combining both change-capturing and smoothing capabilities.

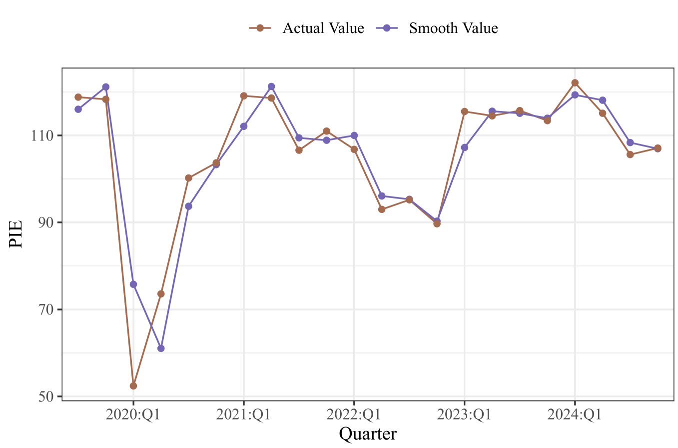

Figure 1. Comparison of PIE quarterly smoothing estimated series with true values. The comparison period spans from the third quarter of 2019 to the fourth quarter of 2024, with PIE being compiled and published since the third quarter of 2019.

Furthermore, since the PIE was not officially recorded prior to the third quarter of 2019, additional analysis was conducted to assess the validity of the forward-smoothed estimates generated by the model. To achieve this, all forward-smoothed estimates of PIE starting from the first quarter of 2017 were compared with indicators closely related to this index. The NBS had previously published the Investment Prosperity Index for the CSE industry, which serves as an indicator of PIE. This index effectively reflects enterprises' confidence in the market environment and their investment intentions, demonstrating a strong correlation with PIE. Additionally, GDP, as one of the most critical indicators for measuring a country's economic development, provides valuable insights into overall macroeconomic performance and exhibits a notable correlation with PIE (r = 0.7057, p < 0.05). As shown in **Figure 2**, the forward-smoothed estimates of PIE exhibit a degree of similarity in trend with both the Investment Prosperity Index and GDP, further indicating that the forward-smoothed estimates hold meaningful reference value.

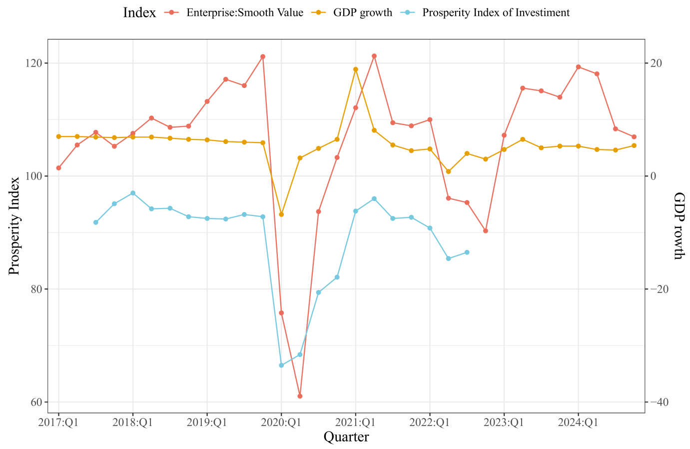

Figure 2. Comparison of PIE with other relevant variables.The comparison period spans from the first quarter of 2017 to the fourth quarter of 2024. The Investment Prosperity Index of CSE industry has been published since the third quarter of 2017 and has not been publicly released since the third quarter of 20.

3.2. Model comparison

This study employed the Root Mean Square Forecast Error (RMSFE) to evaluate the predictive performance of the model:

$$RMSFE=\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{y}_t\text{-}y_t)^2}$$

Where denotes the sample period length, $\widehat{y}_t$ represents the predicted value, and y_t denotes the actual value. To comprehensively evaluate the model's forecasting performance, four benchmark models were selected for comparison: the ARIMA model, two MIDAS models, and the MF-DFM single block model.

3.2.1. Comparison Models 1 and 2: MIDAS-ExAlmon and MIDAS-Beta

The MIDAS models were constructed using three variables: the cumulative growth rate of CSE fixed asset investment, the cumulative growth rate of CSE private fixed asset investment, and the cumulative growth rate of CSE general public budget expenditure. The data spans from 2019:Q3 to 2024:Q4. To define the weighting functions, a two-parameter exponential Almon polynomial and a two-parameter Beta polynomial were employed, respectively [24,25]. Since the MIDAS model relies on balanced panel data, any missing monthly data points were filled using the most recently published values.

3.2. Comparison Model 3: ARIMA

The ARIMA model was constructed based on PIE data spanning from 2019:Q3 to the quarter preceding the target forecasting period. The optimal parameters are selected using the Akaike Information Criterion, and one-step-ahead rolling forecasts are subsequently performed.

3.2.3. Comparison Model 4: MF-DFM single block model

This model is a simplified version of the MF-DFM used in this study. It employed only a single factor block for predicting model outcomes and did not control for the cultural, sports, and entertainment industries. The rolling forecasting approach followed the same procedure as the complete model.

3.2.4. Model comparison

To objectively reflect the model's fitting performance, we presented the results across four distinct data intervals. Since the PIE of CSE is only available from 2019:Q3 onward, it is not possible to evaluate the model's performance before the COVID-19 pandemic in 2020. As an alternative, the selected four intervals cover the period from the of 2022:Q1 to the 2024:Q4. During this timeframe, China's approach to managing COVID-19 underwent significant changes. From Q1 to Q4 of 2022, China remained under the deep impact of the pandemic, with stringent government control measures limiting population movement, which considerably affected economic activities. On December 26, 2022, the National Health Commission of China announced that, starting from January 8, 2023, COVID-19 infection management would shift to Category B infectious disease control. During 2023:Q1, China's pandemic situation experienced a fundamental shift. Economic activities underwent a rapid transition from stagnation, largely due to widespread infections reducing the population's capacity for economic participation, to a sharp "retaliatory" rebound. Given that PIE is calculated based on subjective survey data with a comparative nature relative to the same period in the previous year, the PIE values published from the first quarter of 2023 to the first quarter of 2024 show a significant surge. Starting from the second quarter of 2024, the data returns to normal fluctuation levels. In summary, the PIE data from 2022:Q1 to 2024:Q4 can be characterized by three distinct phases: the pandemic lockdown phase (from 2022:Q1 to 2022:Q4), the post-reopening fluctuation phase (from 2023:Q1 to 2024:Q1) and the stabilization phase (from 2024:Q2 to 2024:Q4).

Firstly, the predictive performance of the models during the stabilization phase was examined. The results indicated that the MF-DFM demonstrates exceptional forecasting accuracy, outperforming the MF-DFM single block, both MIDAS models, and the ARIMA model. In the prediction of the pandemic lockdown phase, the MF-DFM also exhibited the highest accuracy. Since the number of PIE data points within this period is relatively limited, the stability of all models was somewhat reduced. Nevertheless, under these conditions, the MF-DFM and single block models showed a more pronounced accuracy advantage compared to the other models (**Table 2**).

The model fitting results for all models over the full sample period (from 2022: Q1 to 2024:Q4) showed unsatisfactory performance, with RMSFE values exceeding 10. Among these models, the MF-DFM performed slightly worse than the MIDAS-ExAlmon model, ranking second in predictive accuracy across the five models. This suggested that during the post-reopening fluctuation phase, the MF-DFM struggled to capture the data's volatility effectively. To further investigate, we calculated the models' predictive performance within the full sample period after excluding the fluctuation phase. In this adjusted interval, the MF-DFM demonstrated excellent predictive performance, significantly outperforming both MIDAS models and the ARIMA model.

Table 2. RMSFE comparison of models in different prediction intervals

Model	2024Q2–2024Q4	2022Q1–2022Q4	2022Q1–2024Q4	2022Q1–2024Q4*
MF-DFM	1.98	5.23	10.46	4.16
MF-DFM single block	2.79	5.59	10.56	4.60
ARIMA	10.08	7.08	11.28	8.49
MIDAS-ExAlmon	3.48	10.59	9.77	8.32
MIDAS-Beta	2.20	8.28	14.29	6.42

Note: The bold text indicates the best-performing model in each forecast interval. * indicates that the full sample period excludes the highly volatile period (from the first quarter of 2023 to the first quarter of 2024).

3.2.5. Monthly forecasting

In addition to the cross-model comparison, we also examined the performance of each model in predicting PIE across the three forecasting months within a given quarter (the third month of the previous quarter, and the first and second months of the current quarter). The evaluation considered monthly forecasts starting from December 2021, with results grouped by forecast month. To ensure stability in the comparison, we excluded the highly volatile period from 2023:Q3 to the 2024:Q4. The results indicated that when forecasting PIE for a given quarter, the MF-DFM model's accuracy steadily improved as the forecast month progresses. Furthermore, the MF-DFM consistently outperformed the single block model in terms of predictive accuracy (**Table 3**).

Table 3. RMSFE comparison of predicted month order

Forecast month	MF-DFM	MF-DFM-single block
First month	5.59	5.85
Second month	3.89	4.59
Third month	3.62	3.97

Note: The forecast period spans from the first quarter of 2022 to the fourth quarter of 2024, excluding the first quarter of 2023 to the first quarter of 2024. Bold text indicates the month with the best forecast performance in a vertical comparison.

4. Discussion

This study presents the first real-time forecasting model for macroeconomic indicators in the CSE industry. While the NBS classifies the CSE sector as a distinct statistical category, no scholars have yet developed a real-time forecasting model for this sector, whether based on the standalone industry statistics commonly used abroad or the combined industry statistics adopted in China. This study employs a Mixed-Frequency Dynamic Factor Model (MF-DFM) to construct a real-time forecasting framework, successfully developing a model for predicting the PIE of CSE. Empirical analysis demonstrates that the model effectively captures PIE's trend at the fitting stage, accurately acquiring past information and generating trend-consistent smoothed estimates for the latent monthly PIE values. Given the relatively short time span of the PIE series, this study also investigates the model's forward-smoothing results for periods prior to PIE's official reporting. The findings reveal that the forward-smoothed estimates of PIE closely align with the trend of the Investment Prosperity Index, previously reported by the NBS. Traditional dynamic factor models primarily focus on GDP, the most critical economic variable, and typically do not address backward smoothing. The findings of this study suggest that dynamic factor models can also achieve reasonable accuracy in backward smoothing. This contributes valuable insights for imputing past time series data and converting mixed-frequency data across different time scales.

In the out-of-sample rolling forecast, the MF-DFM demonstrates strong predictive accuracy for PIE during periods of normal fluctuation, outperforming traditional models such as the ARIMA and MIDAS models. This result highlights the notable advantage of the MF-DFM in handling mixed-frequency, high-frequency, and unbalanced time series data. However, when predicting the full sample period that includes the post-reopening fluctuation phase, the MF-DFM exhibits relatively lower accuracy. One possible explanation for this outcome relates to the unique characteristics of PIE, as previously discussed. PIE is a distinctive indicator that incorporates subjective survey responses and year-on-year comparisons. During the post-reopening fluctuation phase, respondents in the business climate survey may have perceived a substantial improvement in the economic environment compared to the same period in the previous year. This subjective perception may have been amplified, resulting in certain distortions in PIE relative to objective macroeconomic indicators. Since MF-DFM captures macroeconomic trends in the CSE industry, the PIE forecasts given will be closer to the real situation without subjective bias.

This study pays particular attention to certain characteristics of the MF-DFM model. We examine how the forecast month within the same quarter influences prediction accuracy. Clearly, when forecasting a given quarter, the accumulation of statistical information and the availability of more observed values to correct the smoothed estimates should enhance prediction accuracy. The complete MF-DFM model accurately reflects this trend, demonstrating particularly strong predictive performance in the month preceding the release of quarterly data. Additionally, the model comparison results indicate that the complete MF-DFM model outperforms the MF-DFM single block model, which does not control for correlations among CSE industry indicators. The complete model effectively estimates the structured development trends of the three industries. For indicators beyond GDP, especially macroeconomic indicators representing individual or combined industries, the dynamic factor model continues to effectively capture valuable information and deliver accurate forecasts.

Unlike the commonly used forecasting variable in dynamic factor models, namely GDP, this model adopts PIE, which incorporates elements of subjective survey data, as the target indicator. The study finds that the smoothed estimates of PIE generally align with the trends of macroeconomic forecasting indicators, indicating that the model successfully captures macroeconomic trend changes. This also demonstrates that PIE is capable of describing macroeconomic dynamics. Given that China has not yet released quarterly CSE GDP growth rates, this indicator holds significant value for enterprises in planning production scales and adjusting business strategies. By providing accurate quarterly data guidance, PIE can help firms mitigate the risk of decision-making errors.

This study has certain limitations. First, the model assumes that PIE exhibits the characteristics of a year-on-year series. During the preliminary modeling stage, we observed that modeling the year-on-year PIE series resulted in poor predictive accuracy, with the forecasts heavily relying on the previous period's value, thereby lacking robustness. However, due to limitations in the CSE dataset, this study is unable to verify whether PIE aligns more closely with a year-on-year or month-on-month series. Since most CSE industry data in China are presented as year-on-year data without published actual values, converting PIE into a month-on-month series is not feasible. Future research may further explore the time series properties of PIE. Second, PIE itself, as the forecast variable, presents certain limitations. On one hand, the PIE series is relatively short, making it difficult to fully ensure numerical stability in the model. On the other hand, although this study indirectly demonstrates that PIE, which reflects subjective perceptions from surveyed groups, aligns with macroeconomic trends, PIE fundamentally differs from traditional macroeconomic indicators. Consequently, relying on PIE predictions as a reference for industry indicators inevitably introduces some degree of human bias. However, given the constraints of China's statistical data release practices, there are currently no more suitable variables available for real-time forecasting of the CSE industry. This underscores the need for improved variable selection, potential development of new indicators, cross-validation with similar indicators in other countries, and the construction of more advanced forecasting models with enhanced functionality. Lastly, while MF-DFM represents one of the most advanced real-time forecasting models available, it also has certain limitations. This study incorporates 26 variables in the model, but not all variables contribute positively to its performance. Although previous studies have demonstrated the value of information decomposition methods in analyzing the contribution of new information the potential contribution of each variable cannot be directly addressed within the current model framework. Future research could explore variable selection mechanisms based on information criteria or machine learning algorithms to eliminate redundant variables, thereby improving the model's operational efficiency and predictive accuracy.

5. Conclusion

This study successfully develops the first real-time forecasting model designed to predict macroeconomic indicators in the CSE industry, selecting PIE as the forecasting variable. The proposed model demonstrates strong predictive accuracy, significantly outperforming traditional econometric models, and effectively captures the macroeconomic dynamics of the CSE industry. The model's forecasting results hold substantial significance for real-time assessment of industry development trends, guiding social production, and optimizing both corporate and individual investment decisions.

Disclosure statement

The author declares no conflict of interest.

References

[1] Yin X, 2005, New Trends of Leisure Consumption in China. Journal of Family and Economic Issues, 26(1): 175.

[2] Yu H, 2014, Between Corporate Development and Public Service: The Cultural System Reform in the Chinese Media Sector. Media, Culture & Society, 36(5): 610–627.

[3] Fang K, 2002, The Development of Industrial Classification Standards of China and Its Comparison with International Standards. Statistical Research, 07: 27–31.

[4] Jin H, 2020, Regional Disparity of Output, TFP and Convergence in Culture, Sports and Entertainment Industry in China. Journal of Nanjing University of Finance and Economics, 06: 46–55.

[5] He D, Ren W, 2019, Research on the Impact of Industrial Policies on the High-Quality Development of China's Cultural Industry. Jiangsu Social Sciences, 01: 19–27.

[6] Liu H, 2024, Unveiling the Economic Potential of Sports Industry in China: A Data Driven Analysis. Plos One, 19(9): e0310131.

[7] Yang K, 2020, The Construction of Sports Culture Industry Growth Forecast Model Based on Big Data. Personal and Ubiquitous Computing, 24(1): 5–17.

[8] Hu J, 2019, ISPO Shares the Business Potential of the Sports Industry. China's Foreign Trade, 1.

[9] Han S, 2012, Empirical Analysis of the Forecasting Effect of Enterprise Prosperity Index and Application Discussion. Modern Business Trade Industry, 24(03): 13–15.

[10] Lehmann R, 2023, The Forecasting Power of the Ifo Business Survey. Journal of Business Cycle Research, 19(1): 43–94.

[11] Yang D, 2022, Exploration of the Use of Key Tax Source Enterprise Data to Compile a Boom Index. Tianjin Economy, 02: 44–49.

[12] Shi F, Hu Y, 2021, Prediction on Chinese Business Climate Index and Entrepreneur Confidence Index. Statistics & Decision, 37(03): 41–45.

[13] Bok B, Caratelli D, Giannone D, et al., 2018, Macroeconomic Nowcasting and Forecasting with Big Data. Annual Review of Economics, 10(1): 615–643.

[14] Higgins P, 2014, GDPNow: A Model for GDP 'Nowcasting'. Federal Reserve Bank of Atlanta, GA.

[15] Wang X, Si N, Song T, 2021, Nowcasting China's Quarterly GDP Using Mixed-Frequency Data. Journal of Financial Research, 494(8): 22–41.

[16] Zheng T, Chen W, Wang X, 2023, Daily Measurement of Economic Uncertainty Based on Mixed Frequency Data and Its Applications. Statistical Research, 40(1): 33–48.

[17] Yi Y, Huang D, Wang X, 2024, Nowcasting GDP with Big Macroeconomic Data. China Economic Quarterly, 24(03): 843–860.

[18] Kabundi A, Mbelu A, 2021, Estimating a Time-Varying Financial Conditions Index for South Africa. Empirical Economics, 60(4): 1817–1844.

[19] Algaba A, Borms S, Boudt K, et al., 2023, Daily News Sentiment and Monthly Surveys: A Mixed-Frequency Dynamic Factor Model for Nowcasting Consumer Confidence. International Journal of Forecasting, 39(1): 266–278.

[20] Doz C, Giannone D, Reichlin L, 2012, A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. Review of Economics and Statistics, 94(4): 1014–1024.

[21] Mariano R, Murasawa Y, 2003, A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. Journal of Applied Econometrics, 18(4): 427–443.

[22] Zheng T, Wang X, 2013, Measuring China's Business Cycle with Mixed-Frequency Data and Its Real-Time Analysis. Economic Research Journal, 48(6): 58–70.

[23] Bańbura M, Modugno M, 2014, Maximum Likelihood Estimation of Factor Models on Datasets With Arbitrary Pattern of Missing Data. Journal of Applied Econometrics, 29(1): 133–160.

[24] Ghysels E, Kvedaras V, Zemlys-Balevičius V, 2020, Mixed Data Sampling (MIDAS) Regression Models. Handbook of Statistics, 42: 1–34.

[25] Kuzin V, Marcellino M, Schumacher C, 2011, MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area. International Journal of Forecasting, 27(2): 529–542.