The Application of Statistical Methods in Economic Forecasting in the Era of Big Data

Li Chen

Linyi City Highway Affairs Development Center, Yishui County Center, Linyi 276400, Shandong, China

Abstract: With the rapid development of information technology, data is being generated and accumulated at an unprecedented scale and speed. The sources of economic operation data have expanded from traditional statistical surveys to multiple channels such as Internet transactions, social media, satellite remote sensing, financial markets and industrial chain logistics. Traditional statistical analysis and economic forecasting methods have become difficult to cope with the massive, unstructured and dynamic characteristics of data. This article unfolds from four aspects: machine learning regression optimization, the fusion of time series statistics and dynamic Bayesian, multi-dimensional factor dimensionality reduction analysis, and the cross-application of statistical learning and text mining, aiming to reveal the transformation logic and practical significance of traditional statistical models in the big data environment. By constructing a data-driven prediction framework, a leap from parametric inference to adaptive learning is achieved, providing technical and systematic theoretical references for economic trend analysis and macro decision-making.

Keywords**:** big data; statistical methods; economic forecast

1. Introduction

Since the beginning of the 21st century, the global economic operating environment has become increasingly complex and volatile. The pace of macro-control, financial market and industrial structure adjustment has been continuously accelerating, and the importance of economic forecasting has been pushed to a new height. Meanwhile, the explosive development of information technology has given rise to a fundamental transformation in data forms. The advent of the big data era has made the digital, networked and intelligent characteristics of economic activities increasingly prominent. The sources of economic statistics have extended from traditional surveys and reports to multiple channels such as Internet behavior data, social media public opinions, e-commerce transactions, financial market fluctuations, and even satellite remote sensing images. This not only greatly expands the breadth and depth of economic information but also poses unprecedented challenges to traditional statistical methods. Traditional regression and time series models perform well in environments with limited sample size and single variable structure, but their predictive performance significantly declines in high-dimensional, nonlinear and dynamically evolving economic systems. As a result, how to integrate machine learning, data mining and artificial intelligence technologies within the logical framework of statistical inference has become the core issue of the transformation of statistics[1].

2. New characteristics faced by economic forecasting in the era of big data

In the era of big data, economic forecasting has presented unprecedented complexity and dynamic characteristics. Firstly, the diversification of data sources has fundamentally transformed the way economic information is obtained. Traditional statistical survey data and financial transaction records have gradually been deeply integrated with multi-source data such as Internet search, social media public opinions, e-commerce price indices, and satellite remote sensing images, forming a pattern where structured, semi-structured, and unstructured data coexist. Secondly, the explosive growth of data volume and the demand for real-time performance have made the economic operation state highly dynamic. Prediction models must have the ability to update and iterate rapidly to cope with the impact of market fluctuations and unexpected events. Furthermore, the high dimensionality and heterogeneity of big data lead to more complex relationships among variables, making it difficult for traditional linear models to reveal the nonlinear laws and interaction effects hidden behind the data. Furthermore, the problems of information redundancy and noise interference have become prominent, and the model needs to rely on statistical learning and feature selection techniques for signal extraction and dimensionality reduction processing. The continuous development of the economic system towards digitalization has strengthened the real-time decision-making attribute of predictions, which puts forward requirements for statistical methods. Statistical methods need to have high accuracy, as well as interpretability and adaptability, to meet the demands of policy regulation and market supervision in intelligent decision-making. These emerging features, when combined together, A brand-new technological ecosystem for economic forecasting in the era of big data has been formed.

3. The demand for economic forecasting in the era of big data

3.1 Requirements for real-time prediction and dynamic update capabilities

Traditional economic forecasting usually relies on quarterly or annual statistical data, which is characterized by significant lag and slow response. It is difficult to meet the needs of rapid market and policy decision-making. The current economic system shows a high degree of interactivity and complexity. Financial fluctuations, international trade data or changes in public opinion in a country may all cause significant fluctuations in macroeconomic indicators within a short period of time. The predictive model must have the ability to update in real time and self-calibrate dynamically. It can immediately adjust parameters and structure when new data streams are input, achieving continuous tracking of economic trends. For instance, by conducting real-time monitoring of high-frequency trading data, electronic invoice traffic, and Internet consumption indices, it can dynamically capture short-term changes in economic activities[2].

3.2 Demand for the integration and analysis of multi-source heterogeneous data

Nowadays, economic forecasting is no longer confined to macro-statistical data and is beginning to show a trend of multi-source and heterogeneous development. Data generated by Internet platforms, financial systems, logistics networks, meteorological monitoring, and social media come in various forms, including structured numerical information, semi-structured and unstructured text, images, and semantic data. This situation requires statistical methods to achieve cross-platform and cross-type data integration and cleaning, extracting the core signals of economic activities from noise and redundant information.

3.3 Requirements for the interpretability of prediction results and policy usability

In the process of economic forecasting shifting from technical analysis to decision support, the interpretability of models has become a key measurement criterion. Although big data models have made certain progress in terms of accuracy, the "black box" trend they present has led to a lack of transparency in the prediction results, which is quite unfavorable for government departments and research institutions to carry out policy verification and risk assessment. Economic forecasts need to answer the question of "how the future will change" and also explain "why it will change in this way". When constructing models, statistical methods should focus on the economic significance, causal structure and traceability of variables to ensure that the prediction conclusions have policy reference significance.

4. Innovative Application of Statistical Methods in Big Data Economic Forecasting

4.1 Statistical Regression Optimization Based on Machine Learning

In the big data environment, traditional multiple linear regression models have become increasingly difficult to comprehensively present the nonlinear relationships among economic variables. After the introduction of machine learning methods, statistical modeling has acquired the capabilities of self-learning and pattern recognition. Algorithms such as random forests, gradient boosting trees, and support vector regression, By automatically selecting key variables and identifying nonlinear mapping relationships in high-dimensional Spaces, the traditional model's reliance on functional forms has been broken through. In actual economic forecasting work, researchers often combine statistical regression and machine learning to form hybrid models. For instance, when predicting inflation, they incorporate energy prices, consumer search indices, and international commodity fluctuations data into input variables. After screening through LASSO or Ridge regression, they are then handed over to GBDT for modeling. This way, both the interpretability of the model and the prediction accuracy can be taken into account. The core innovation of this type of method lies in taking statistical inference as the theoretical support and algorithm optimization as the implementation approach, achieving a methodological transformation from "empirical assumption" to "data-driven"[3].

4.2 Integration of Time Series Statistics and Dynamic Bayesian Inference

The fluctuations of the economic system exhibit very prominent time-varying and non-stationarity characteristics. It is difficult to accurately depict its dynamic evolution features by relying solely on a single static model. In the big data environment, the innovation of time series statistical methods lies in their mutual integration with dynamic Bayesian models and hidden Markov models. This type of method introduces the mechanisms of state transition and probability update in the time dimension, enabling parameter estimation to be adjusted in real time as new data continuously flows in. Taking the domestic PMI index prediction as an example, this model can combine historical macro indicators and real-time network transaction volume, and dynamically correct the parameter distribution with the help of Bayesian filtering, effectively improving the timeliness and robustness of the prediction. Compared with the traditional ARIMA model, dynamic Bayesian inference can simultaneously take into account the uncertainty of data and external shock effects. It demonstrates more outstanding stability and explanatory power in the process of high-frequency economic data analysis[4].

4.3 Dimensional Reduction Innovation of Multi-dimensional Factor and Principal Component Analysis

In the face of economic big data, there are tens of thousands of indicator variables. If direct modeling is carried out, the computational cost will be extremely high, and the results will also be distorted due to multicollinearity. Under such a background, principal component analysis (PCA) and factor analysis (FA) in statistics have achieved expansion and innovation in methods. On the one hand, Some researchers have combined traditional linear dimensionality reduction with sparse coding, independent component analysis (ICA), and other algorithms, enhancing the stability of factor extraction. On the other hand, the statistical covariance matrix estimation and eigenvalue decomposition methods have been strengthened, enabling the model to accurately identify the main economic drivers in extremely high-dimensional data. For instance, when constructing the "Comprehensive Prosperity Index" of the macroeconomy, the model can compress multiple dimensions of indicators such as price levels, credit scales, employment data, and online consumption into several principal components to capture the common fluctuations in economic prosperity and achieve a quantitative description of macro trends.

4.4 Integrated Application of Statistical Learning and Text Mining

The addition of unstructured text data has greatly enriched the information sources for economic forecasting. Traditional statistical methods mainly rely on manual encoding or word frequency analysis when dealing with text data. However, in the context of big data, the combination of statistical learning and natural language processing (NLP) technology enables text information to enter the prediction model in a quantitative way. Through statistical text mining tools such as TF-IDF, Word2Vec and sentiment tendency analysis, implicit variables such as "economic confidence" and "market sentiment" can be extracted from news reports, social media and policy documents. For instance, researchers utilized the financial public opinion on Weibo to construct a market sentiment index and combined it with the stock market volatility model (GARCH), significantly enhancing the accuracy of short-term volatility prediction. Such cross-border applications not only broaden the applicable boundaries of statistical methods but also drive economic forecasting from structured data analysis to an intelligent stage of semantic information fusion[5].

5. Conclusion

Overall, the advent of the big data era has driven a systematic innovation in economic forecasting paradigms, and the role of statistical methods has shifted from traditional empirical modeling to data-driven intelligent inference. Statistics, with its rigorous logical and mathematical foundation, provides a framework for structured interpretation of complex economic data. The integration of new technologies such as machine learning and text mining endows it with the ability to adapt dynamically and characterize nonlinearly. Economic forecasting has thus moved from static sampling analysis to a stage of real-time feedback and multi-source integration, becoming an important bridge connecting macro policies, market mechanisms and social public opinions.

References

[1] Fu Yu. Research on the Innovative Development of Haikou's Trade and Logistics Industry under Big Data Technology [J Logistics Science and Technology, 2020,48(21):59-61.

[2] Huang Yuning. Data Element Allocation, Technological Innovation and High-Quality Development of Low-altitude Economy [J]. Industrial Technology Economics,2025,44(11):43-52.

[3] Du Yanxin. Research on Public Sports Service Demand Prediction Model Based on Big Data [J] Stationery, Sports Goods and Technology,2025,(21):13-15.

[4] Li Zhen. Application of Statistical Methods in Economic Forecasting in the Era of Big Data [J]. Business 2.0,2025,(24):58-60.

[5] Lu Hanlin. The Application of Big Data Statistical Methods in Economic Forecasting [J]. Today's Wealth (China Intellectual Property),2022,(11):64-66.

Author Bio

Li Chen (1974-), female, Han ethnicity, native of Yishui, Shandong Province, holds a bachelor's degree. Research focus: Economic statistics.