AEM JOURNAL .ORG

Applications of High-Dimensional Factors and Machine Learning in Behavioral Macroeconomic Forecasting

Zimo Fei[1,*]
[1]University of California, San Diego (UCSD), La Jolla, CA, 92093
Corresponding author: Zimo Fei (email: xxcngl@163.com).

ABSTRACT This paper develops a unified macroeconomic forecasting framework that integrates high-dimensional factors, machine learning methods, and behavioral variables, and systematically examines their joint performance in forecasting growth and inflation. First, common high-dimensional factors are extracted from a large panel of macroeconomic indicators in order to compress the shared information contained in data-rich environments. Second, behavioral information, including consumer confidence, expectation measures, sentiment indicators, and uncertainty proxies, is incorporated into the forecasting system to capture the role of beliefs, perceptions, and sentiment in macroeconomic fluctuations. Third, the out-of-sample forecasting performance of benchmark time-series models, factor-based models, and machine-learning-based models is compared within a unified framework. The results show that high-dimensional factors significantly improve upon traditional benchmark models, behavioral variables provide additional forward-looking information beyond common latent factors, and the hybrid framework combining high-dimensional factors, behavioral variables, and machine learning delivers the strongest forecasting performance. Further analysis indicates that the predictive contribution of behavioral variables is strongly state-dependent and becomes especially pronounced at short horizons and during periods of elevated uncertainty. The findings suggest that improvements in macroeconomic forecasting depend not only on high-dimensional information compression and nonlinear modeling capacity, but also on the effective incorporation of behavioral signals related to expectations, sentiment, and uncertainty. This study provides new evidence on the empirical relevance of behavioral macroeconomics for forecasting and offers useful implications for future macroeconomic monitoring and policy analysis.

INDEX TERMS High-dimensional factors, machine learning, behavioral variables, macroeconomic forecasting

I. INTRODUCTION

IN an environment characterized by growing economic complexity, frequent external shocks, and increasingly rapid information transmission, macroeconomic forecasting faces challenges of unprecedented scale [1]–[3]. On the one hand, macroeconomic dynamics are jointly shaped by a wide range of factors, including output, employment, prices, monetary conditions, financial markets, and external disturbances, making traditional forecasting frameworks based on a limited set of variables and linear relationships increasingly inadequate for capturing the complexity of the economic system [4], [5]. On the other hand, behavioral forces such as household and firm expectations, market sentiment, policy uncertainty, and media narratives have become ever more influential in shaping macroeconomic fluctuations [6], [7]. As a result, macroeconomic dynamics are no longer driven solely by fundamentals, but also by the evolving interaction of beliefs, perceptions, and sentiment [8]. Against this background, a central question in contemporary macroeconomic research is how to integrate traditional macroeconomic information with behavioral signals in a high-dimensional data environment so as to improve both forecasting accuracy and economic interpretability.

The existing literature has advanced macroeconomic forecasting along two main lines. The first is the development of high-dimensional factor models, which extract a small number of common factors from a large panel of macroeconomic indicators and thereby mitigate information redundancy and multicollinearity in data-rich environments [9], [10]. The second is the growing application of machine learning methods to macroeconomic forecasting. Compared with conventional linear models, machine learning approaches offer greater flexibility in handling variable selection, nonlinear relationships, interaction effects, and structural change. At the same time, the behavioral macroeconomics literature highlights that bounded rationality, expectation biases, sentiment fluctuations, and narrative transmission can systematically affect consumption, investment, and price-setting decisions. Accordingly, variables such as consumer confidence, survey expectations, news sentiment, and policy uncertainty may contain predictive content beyond that embedded in conventional macroeconomic indicators. However, the existing literature has tended to examine high-dimensional factors, machine learning techniques, and behavioral variables in isolation, with relatively limited attention paid to their potential complementarity within a unified forecasting framework.

Motivated by this gap, this paper develops an integrated macroeconomic forecasting framework that combines high-dimensional factors, machine learning methods, and behavioral variables. The central objective is to examine whether behavioral information provides additional out-of-sample predictive gains in a high-dimensional setting, and whether machine learning methods can more effectively capture the complex relationships between latent macroeconomic factors and behavioral signals. Specifically, the paper first extracts common high-dimensional factors from a large set of macroeconomic indicators to summarize the shared information in the economy. It then incorporates behavioral variables, such as consumer confidence, expectation measures, sentiment indicators, and uncertainty indices, into the forecasting system. Finally, it conducts a systematic comparison across benchmark time-series models, factor-based models, and machine-learning-based models in order to evaluate their relative forecasting performance. The contribution of this paper is threefold. First, it systematically introduces behavioral variables into a high-dimensional macroeconomic forecasting framework, thereby extending the empirical application of behavioral macroeconomics. Second, it clarifies the complementary roles of high-dimensional factors and machine learning in processing behavioral information. Third, through out-of-sample forecasting comparisons and variable-importance analysis, it provides new empirical evidence on the role of behavioral factors in macroeconomic fluctuations and forecasting performance.

II. LITERATURE REVIEW

A. HIGH-DIMENSIONAL FACTORS IN MACROECONOMIC FORECASTING

The development of high-dimensional factor models has provided an important methodological foundation for macroeconomic forecasting [11]. In traditional forecasting frameworks, researchers typically rely on a small set of key variables within linear specifications. Such approaches, however, become increasingly restrictive when the available information set expands substantially. Macroeconomic indicators are often highly correlated and informationally redundant, and the rapid growth in data availability makes it difficult for low-dimensional models to adequately capture broad-based economic movements. In this context, the diffusion index approach developed by Stock and Watson offers a canonical solution for forecasting in data-rich environments. Their central idea is to extract a small number of common factors from a large panel of time series using principal component methods and then employ these latent factors to forecast key macroeconomic outcomes. The importance of this framework lies in showing that even when the number of predictors is very large, the core information in the economy can still be summarized parsimoniously through a low-dimensional latent structure.

Building on this line of research, Bai and Ng addressed one of the central econometric issues in approximate factor models, namely, how to determine the number of relevant factors [12]. Their contribution is particularly important because the identification of factor dimensionality directly affects both the quality of information compression and the performance of subsequent forecasting models. By proposing information criteria that consistently estimate the number of factors, they provided a rigorous econometric basis for the empirical implementation of high-dimensional factor methods. As a result, the factor-model literature moved beyond a purely heuristic dimension-reduction strategy and developed into a more disciplined framework for macroeconomic analysis.

Taken together, the literature on high-dimensional factors has established a broad consensus that common factors can serve as effective statistical summaries of large macroeconomic information sets and can improve the forecasting of key variables such as output growth and inflation [13]. For this reason, factor models have become an important bridge between traditional macroeconomic forecasting and modern high-dimensional data analysis. At the same time, however, most factor-based forecasting models remain fundamentally linear in structure, which limits their ability to accommodate nonlinearities, state dependence, and complex interaction effects. This limitation creates a natural opening for the incorporation of machine learning methods.

B. MACHINE LEARNING IN DATA-RICH MACROECONOMIC FORECASTING

As the scale, frequency, and diversity of macroeconomic data continue to expand, machine learning methods have increasingly been incorporated into macroeconomic forecasting in order to address several limitations of conventional linear models [14]. Relative to standard regression-based approaches, machine learning methods are more flexible in handling variable selection, functional-form uncertainty, and complex dependence structures. This makes them particularly well suited to high-dimensional, sparse, nonlinear, and interaction-rich data environments. In the context of macroeconomic forecasting, such flexibility implies that machine learning may not only make fuller use of large information sets, but also better capture threshold effects, regime dependence, and structural changes in macroeconomic relationships [15].

Existing evidence suggests that machine learning holds considerable promise in data-rich macroeconomic forecasting. Using U.S. inflation forecasting as a leading example, Medeiros et al. systematically compare several machine learning approaches with conventional benchmark models and show that, in the presence of a large set of predictors, machine learning models can outperform traditional benchmarks on an out-of-sample basis. In particular, their results highlight the strong performance of random forests. The broader implication of this literature is that high-dimensional information sets do not necessarily lead to overfitting, provided that appropriate algorithms and rigorous out-of-sample evaluation strategies are employed. Under such conditions, machine learning can extract stable predictive signals from complex macroeconomic data.

Importantly, the rise of machine learning does not imply that earlier approaches should be discarded. Instead, a growing body of work suggests that machine learning is best viewed as complementary to factor models and conventional time-series methods. High-dimensional factor models are particularly useful for compressing broad-based macroeconomic information, whereas machine learning techniques are better suited to identifying complex mappings between latent factors, raw predictors, and forecasting targets. Consequently, a promising research agenda lies not in treating factor models and machine learning as competing alternatives, but in examining how they can be combined within a unified forecasting framework. This perspective provides a direct methodological motivation for the present study.

C. BEHAVIORAL VARIABLES AND BEHAVIORAL MACROECONOMIC FORECASTING

Whereas high-dimensional factors and machine learning primarily advance macroeconomic forecasting at the methodological level, behavioral macroeconomics highlights an important source of information that conventional forecasting frameworks may fail to capture. The behavioral macroeconomics literature emphasizes that economic agents do not always form expectations under full rationality and perfect information. Instead, bounded rationality, heuristic-based decision rules, belief distortions, sentiment fluctuations, and animal spirits can systematically influence consumption, investment, price-setting, and portfolio choices. In particular, De Grauwe's behavioral macroeconomic framework argues that macroeconomic fluctuations depend not only on fundamental shocks, but also on agents' subjective assessments of the future and their switching across forecasting heuristics. From this perspective, expectations, confidence, sentiment, and narratives are not merely by-products of macroeconomic developments; they may themselves constitute important transmission channels of aggregate fluctuations.

In empirical work, behavioral information is typically operationalized through observable indicators such as consumer confidence, survey expectations, news sentiment, economic policy uncertainty, and central bank communication. These variables are particularly relevant because they may reflect shifts in expectations and risk perceptions earlier than conventional macroeconomic statistics do. Economic policy uncertainty provides a prominent example. Baker, Bloom, and Davis construct a newspaper-based index of policy-related economic uncertainty and show that uncertainty shocks are closely linked to macroeconomic outcomes such as output and employment. Their findings underscore that behaviorally informed text-based indicators can carry economically meaningful predictive content.

More recently, several studies have moved beyond theoretical arguments and directly examined the incremental forecasting value of behavioral variables. Tilly, Ebner, and Livan use emotional and narrative information extracted from global news text to forecast industrial production and consumer prices, finding that sentiment variables significantly improve macroeconomic forecasting performance and that different emotional dimensions exhibit distinct predictive relevance. Similarly, more recent work has started to investigate whether central bank narratives and text-based sentiment indicators can enhance forecasts of macroeconomic variables, suggesting that behaviorally informed signals extracted through machine learning and textual analysis may represent an important new source of information for forecasting growth and inflation. Taken together, this literature indicates that the role of behavioral variables in macroeconomic forecasting has evolved from a theoretical possibility into an empirically testable proposition.

Nevertheless, these variables have rarely been integrated systematically with high-dimensional factor compression and machine-learning-based forecasting within a single framework.

D. LITERATURE GAP AND POSITIONING OF THIS STUDY

In sum, the existing literature has advanced macroeconomic forecasting along three related but largely separate lines: high-dimensional information compression, machine-learning-based prediction, and the introduction of behavioral variables. High-dimensional factor models address the problem of information redundancy in data-rich environments; machine learning methods strengthen the ability to detect nonlinear and complex relationships; and behavioral macroeconomics identifies expectations, sentiment, and narratives as potentially independent drivers of macroeconomic dynamics. Yet these strands of research have mostly developed in parallel. Some studies focus on the predictive efficiency of factor models, others emphasize the technical gains from machine learning relative to traditional benchmarks, and still others test the incremental value of specific behavioral indicators in isolation. By contrast, there remains limited work on how high-dimensional factors and behavioral variables jointly form the information set for forecasting, and how machine learning can be used to identify their combined effects on macroeconomic outcomes within an integrated framework.

To address this gap, the present study develops a unified macroeconomic forecasting framework that combines high-dimensional factors, machine learning, and behavioral variables. The underlying premise is that high-dimensional factors summarize broad macroeconomic comovements, behavioral variables add nontraditional signals related to expectations, sentiment, and uncertainty, and machine learning methods can more effectively capture the complex relationships between these inputs and macroeconomic forecasting targets. Accordingly, this study is concerned not only with whether behavioral variables generate additional out-of-sample predictive gains, but also with whether such gains depend on the front-end compression of macroeconomic information through latent factors and the back-end nonlinear mapping enabled by machine learning methods. It is this integrative perspective that constitutes the main point of departure from the existing literature.

III. THEORETICAL MECHANISM AND RESEARCH HYPOTHESES

A. THEORETICAL MECHANISM

The core theoretical logic of this study rests on three interconnected layers: high-dimensional information compression, the supplementation of behavioral signals, and nonlinear predictive mapping. First, macroeconomic forecasting typically involves an information set that is high-dimensional, structurally complex, and highly correlated. Traditional forecasting approaches often rely on a small number of variables and pre-specified linear relationships, making them ill-suited to the effective use of large-scale information. High-dimensional factor models offer a foundational solution to this problem. Their central idea is to extract a small number of common factors from a large set of macroeconomic indicators in order to summarize broad-based comovements within the economy. The diffusion index framework developed by Stock and Watson shows that this form of factor-based information compression can improve forecasting efficiency in environments with a very large number of predictors. In the present study, high-dimensional factors therefore serve as a front-end device for extracting common macroeconomic information and for constructing a more compact and representative set of state variables for subsequent forecasting models.

However, relying solely on the common macroeconomic information summarized by latent factors may still leave out signals that are forward-looking in nature but are not easily captured by conventional statistical indicators. Behavioral macroeconomics suggests that aggregate fluctuations are not driven exclusively by changes in fundamentals; they are also shaped by expectation formation, sentiment dynamics, belief revisions, and narrative transmission among economic agents. Bounded rationality, heuristic-based decision making, and animal spirits can affect consumption, investment, and price-setting behavior, implying that variables such as consumer confidence, survey expectations, news sentiment, economic policy uncertainty, and central bank communication may carry predictive content. Relative to traditional macroeconomic indicators, these behavioral variables often reflect changes in perceptions of future conditions, risks, and policy environments at an earlier stage. Existing evidence further shows that sentiment- and narrative-based indicators extracted from news text can improve forecasts of industrial production and prices, suggesting that behavioral signals are not mere noise, but potentially valuable complements to conventional macroeconomic information.

On this basis, machine learning provides the key technical mechanism for linking high-dimensional factors, behavioral variables, and macroeconomic forecasting targets. Although conventional linear models are attractive for their simplicity and interpretability, they usually impose relatively stable and additive structures on the data-generating process. In reality, macroeconomic relationships often involve nonlinearities, threshold effects, and state dependence. For instance, the forecasting relevance of behavioral variables may be limited during tranquil periods, but may become substantially stronger during episodes of heightened uncertainty, increased market volatility, or major shifts in expectations. Machine learning methods are better equipped to perform variable selection, approximate complex functions, and identify interaction effects, making them especially suitable for uncovering the complementary roles of latent macroeconomic factors and behavioral signals. Empirical studies have shown that, in data-rich environments, machine learning models can outperform conventional benchmarks under rigorous out-of-sample evaluation.

Accordingly, the general mechanism proposed in this paper can be summarized as follows: high-dimensional factors compress common macroeconomic information, behavioral variables supplement that information with signals related to expectations, sentiment, and uncertainty, and machine learning methods identify the complex mapping from these inputs to macroeconomic forecasting targets. This mechanism forms the theoretical basis for the empirical analysis that follows.

B. RESEARCH HYPOTHESES

Based on the foregoing theoretical mechanism, this paper proposes the following hypotheses.

H1: High-dimensional factors significantly improve macroeconomic forecasting accuracy.

Because the macroeconomy is characterized by a large number of interrelated variables, relying on only a limited set of individual indicators can lead to substantial information loss. By extracting common components from a large information set, high-dimensional factors help alleviate redundancy and multicollinearity and improve forecasting efficiency. Therefore, relative to traditional univariate or low-dimensional linear models, forecasting models that incorporate common high-dimensional factors are expected to exhibit superior out-of-sample predictive performance. The diffusion index literature provides direct support for this view by showing that a small number of factors can summarize the essential predictive content of very large macroeconomic datasets.

H2: Relative to conventional linear models, machine learning methods make more effective use of high-dimensional information and improve forecasting performance.

In high-dimensional settings, the relationship between predictors and macroeconomic targets may not be linear, stable, or additive; instead, it may involve complex interactions and nonlinear mappings. Although factor models effectively reduce dimensionality, their forecasting stage often remains linear in practice. By contrast, machine learning methods offer greater flexibility in handling variable selection, nonlinear effects, and structural instability, and are therefore more likely to identify stable and informative predictive patterns in macroeconomic data. Existing studies on inflation forecasting and other macroeconomic time series suggest that machine learning methods can outperform traditional benchmark models in such environments.

H3: Behavioral variables provide additional out-of-sample predictive gains beyond those captured by high-dimensional factors.

High-dimensional factors mainly summarize the common statistical variation in the macroeconomy, but they do not necessarily capture agents' subjective expectations, sentiment shifts, or perceptions of risk in a complete manner. Behavioral macroeconomics argues that expectations, confidence, sentiment, and narratives are themselves important drivers of macroeconomic fluctuations, implying that corresponding behavioral variables may contain information not fully embedded in conventional macro indicators or latent factor structures. If this is the case, then even after common high-dimensional factors have been incorporated into the model, the inclusion of behavioral variables such as consumer confidence, survey expectations, news sentiment, or policy uncertainty should still produce additional predictive improvements.

H4: The predictive contribution of behavioral variables is state-dependent and becomes stronger during periods of heightened uncertainty.

The role of behavioral variables is unlikely to be constant over time. During relatively stable periods, behavioral information may overlap substantially with fundamentals and thus provide only limited marginal predictive value. During episodes of financial stress, frequent policy adjustment, or elevated macroeconomic uncertainty, however, changes in sentiment, confidence, and narratives are more likely to amplify real economic fluctuations and thereby enhance the forecasting usefulness of behavioral variables. This implies that the incremental contribution of behavioral signals to macroeconomic forecasting should be state-dependent and especially pronounced in high-volatility or high-uncertainty periods. This hypothesis is also consistent with the broader behavioral macroeconomic view that fluctuations in beliefs and animal spirits become more influential in unstable environments.

C. ANALYTICAL IMPLICATIONS FOR THE EMPIRICAL DESIGN

The theoretical mechanism and hypotheses outlined above provide a direct foundation for the empirical design of this paper. First, H1 implies the need to compare benchmark forecasting models with factor-based models in order to identify the baseline predictive value of common latent factors. Second, H2 requires the inclusion of machine learning methods in a unified comparison framework to evaluate whether they can absorb high-dimensional information more effectively than linear approaches. Third, H3 implies that behavioral variables should be incorporated into both factor-based and machine-learning-based forecasting models to test whether they deliver additional out-of-sample predictive gains. Finally, H4 requires the empirical analysis to examine whether the predictive relevance of behavioral variables varies across macroeconomic regimes or periods of different uncertainty levels. In this sense, the empirical strategy of the paper is not simply a mechanical combination of variables and methods; rather, it is organized around the internal logic of information compression, behavioral supplementation, and nonlinear identification, thereby ensuring coherence between the theoretical argument, the research hypotheses, and the empirical tests.

IV. RESEARCH DESIGN

A. FORECASTING TARGETS AND SAMPLE CONSTRUCTION

The empirical analysis of this study is designed to evaluate the joint performance of high-dimensional factors, machine learning methods, and behavioral variables in macroeconomic forecasting. A first step, therefore, is to define the forecasting targets and the sample construction strategy. In line with the existing literature and data availability, the analysis focuses on two representative categories of macroeconomic variables. The first consists of indicators of real economic activity, such as industrial production growth or GDP growth. The second consists of price-related indicators, most notably inflation as measured by the consumer price index. This choice is motivated by two considerations. First, output and inflation are the central targets in macroeconomic monitoring and policy analysis, and they are also the most common forecasting objects in both the factor-model literature and the machine-learning forecasting literature. Second, behavioral variables may affect growth and inflation through different channels, which allows the present study to explore the heterogeneous predictive role of behavioral information across macroeconomic dimensions. The diffusion index literature and related forecasting studies both place output and price dynamics at the center of the analysis.

Regarding data frequency, this study gives priority to monthly observations. Monthly data retain sufficient macroeconomic dynamics while making it easier to integrate relatively high-frequency behavioral indicators such as consumer confidence, survey expectations, news sentiment, and policy uncertainty into a unified forecasting framework. For quarterly variables such as GDP, mixed-frequency treatment may be adopted where necessary, or monthly proxies for real activity, such as industrial production, may be used in order to preserve comparability across models. The broader literature on forecasting in data-rich environments commonly relies on large monthly panels of macroeconomic indicators, while behavioral variables are also typically aggregated at the monthly frequency before entering forecasting models.

As for the sample period, the study is designed to cover a sufficiently long time span that includes multiple business-cycle phases and episodes of varying uncertainty. This is important for ensuring that the evaluation of forecasting performance is not driven by a single macroeconomic regime. The full sample will be divided into training, validation, and testing segments, or alternatively, a rolling-window or expanding-window recursive forecasting design will be implemented. Such procedures help avoid drawing conclusions solely from in-sample fit and are consistent with the forecasting literature's emphasis on rigorous out-of-sample evaluation. This is especially important in machine learning applications, because model comparison and hyper-parameter tuning must be based on validation performance rather than in-sample explanatory power.

B. VARIABLE SYSTEM AND DATA STRUCTURE

To evaluate forecasting performance under different information sets in a systematic manner, all explanatory variables are divided into three groups: conventional macroeconomic variables, high-dimensional factor variables, and behavioral variables. This classification corresponds directly to the core logic of the paper. Conventional macroeconomic variables capture standard fundamentals; high-dimensional factors summarize the common components embedded in a large information set; and behavioral variables provide additional signals related to expectations, sentiment, and uncertainty.

The first group consists of conventional macroeconomic variables. These include standard indicators of output, labor market conditions, prices, money, interest rates, financial markets, and the external sector. For example, real activity variables may include industrial production, retail sales, fixed investment, and unemployment; price variables may include CPI, PPI, and core inflation; and financial and monetary variables may include short-term interest rates, long-term interest rates, money supply, credit aggregates, stock market indices, and exchange rates. This group forms the main body of the high-dimensional macroeconomic information set and serves as the primary source for factor extraction. The factor-model literature is fundamentally based on extracting common latent components from large panels of this kind in order to summarize the general state of the economy.

The second group consists of high-dimensional factor variables. Rather than directly feeding all macroeconomic variables into the forecasting model, this study first extracts a small number of latent common factors from the high-dimensional macroeconomic dataset using principal component analysis or an approximate dynamic factor approach. The number of retained factors will be determined using the information criteria proposed by Bai and Ng, thereby avoiding arbitrary researcher choice. This procedure offers two advantages. First, it substantially reduces dimensionality and alleviates multicollinearity. Second, it preserves the dominant common information contained in a large number of macroeconomic time series. Bai and Ng show that, in large-dimensional panels, the number of factors can be consistently estimated using information criteria, which provides a rigorous econometric foundation for this approach.

The third group consists of behavioral variables. A central contribution of this paper is to operationalize the key elements of behavioral macroeconomics as a set of observable indicators. Specifically, the behavioral block includes confidence-based variables such as consumer confidence and business sentiment; expectation-based variables such as inflation expectations, economic outlook indicators, or professional forecaster surveys; sentiment- and narrative-based variables such as news sentiment, media narrative intensity, or the tone of central bank communication; and uncertainty-based variables such as economic policy uncertainty. These variables are chosen because behavioral macroeconomics emphasizes that bounded rationality, animal spirits, and heuristic expectations can influence macroeconomic fluctuations, while empirical studies show that textual context, narratives, and emotions can improve forecasts of industrial production and price dynamics.

Before estimation, all time series will undergo a unified preprocessing procedure. Depending on their statistical properties, raw variables will be transformed into year-on-year growth rates, month-on-month growth rates, log differences, or first differences in order to address nonstationarity. All explanatory variables will then be standardized prior to factor extraction and machine learning estimation to ensure comparability across variables measured in different units. Missing values may be handled through interpolation, carry-forward rules, or synchronized sample trimming, depending on the structure of the underlying dataset. Extreme values may be controlled through winsorization or robust scaling. Since both factor models and machine learning algorithms are data-driven, consistent data cleaning and transformation are essential for meaningful model comparison.

C. EXTRACTION OF HIGH-DIMENSIONAL FACTORS

The extraction of high-dimensional factors is a key step in the empirical design. This study first constructs a large monthly panel of macroeconomic time series and, after stationarity adjustments and standardization, applies principal component analysis to extract common factors. Principal component methods are widely used in the literature because they are effective in identifying the main directions of variation in large time-series panels and are operationally robust. In the diffusion index framework, Stock and Watson use precisely this approach to compress many predictors into a few latent factors and then use those factors to forecast macroeconomic outcomes.

Formally, let the transformed macroeconomic data matrix be denoted by X_t, with cross-sectional dimension N and time dimension T. The study extracts the first r common factors F_t through principal component decomposition, where the optimal number of factors is selected using the Bai–Ng information criteria. Compared with fixing the number of factors a priori, this data-driven procedure reduces arbitrariness and improves reproducibility. Since different factors may loosely correspond to latent dimensions such as real activity, price dynamics, financial conditions, or external shocks, the empirical analysis may also provide descriptive interpretations of the correlation structure, although such interpretations are not imposed as identifying assumptions.

To identify whether behavioral variables contain information beyond the common macroeconomic component, behavioral indicators will not be included in the factor extraction stage. This design choice is crucial. The aim of the paper is not to compress all available variables mechanically into a single low-dimensional representation, but rather to distinguish between common macroeconomic information and behaviorally informed signals. If behavioral variables were incorporated directly into the factor extraction procedure, the subsequent marginal contribution of behavioral information would become more difficult to identify. The strategy adopted here is therefore to extract latent factors only from the conventional high-dimensional information set and then combine these factors with behavioral variables as separate inputs in the subsequent forecasting models.

D. MODEL SPECIFICATION

To assess the relative roles of high-dimensional factors, machine learning, and behavioral variables in macroeconomic forecasting, the empirical analysis will construct four increasingly rich classes of forecasting models.

The first class consists of traditional benchmark models. These may take the form of autoregressive models or simple linear regressions with a limited number of standard controls. Their role is not to generate the best possible forecasts, but to provide a transparent baseline against which the incremental value of factors and behavioral variables can be evaluated.

The second class consists of factor-augmented models. These models add the extracted common factors from the high-dimensional macroeconomic dataset to the benchmark specification, thereby forming a factor-augmented forecasting framework. A standard specification can be written as:

$$y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{k=1}^{r} \gamma_k F_{k,t} + \varepsilon_{t+h}, \tag{1}$$

where y_{t+h} denotes the forecasting target at horizon h, and $F_{k,t}$ denotes the k-th common factor. This class of models is designed to test whether information compression through latent factors improves forecasting performance. The diffusion index literature provides the main methodological foundation for this specification.

The third class consists of behaviorally augmented linear models. These extend the factor-augmented specification by incorporating a vector of behavioral variables B_t:

$$y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{k=1}^{r} \gamma_k F_{k,t} + \sum_{m=1}^{q} \delta_m B_{m,t} + \varepsilon_{t+h}. \tag{2}$$

The purpose of this model is to test whether behavioral variables continue to provide predictive gains even after common macroeconomic factors have been accounted for, corresponding to H3. If this model outperforms the pure factor model out of sample, then the behavioral variables can be interpreted as containing predictive content not fully absorbed by common macroeconomic factors. Evidence from news-sentiment forecasting provides direct motivation for such a specification.

The fourth class consists of machine learning models and hybrid models. To capture nonlinearities and interaction effects, the study will introduce several representative machine learning algorithms, such as Lasso, random forests, and gradient boosting methods. In implementation, machine learning models will be trained under multiple input configurations: using only conventional macroeconomic variables, using only latent factors and lags, and using factors together with behavioral variables. This setup allows for a transparent comparison of how different information sets perform under different forecasting algorithms. The machine-learning inflation forecasting literature shows that, in environments with many predictors, random forests can outperform traditional benchmark models and handle complex predictor structures effectively.

With respect to estimation details, lag lengths in linear models may be chosen based on information criteria or rolling validation performance, while hyperparameters in machine learning models will be tuned on training and validation samples using grid search, cross-validation, or time-series-based rolling validation. Since the purpose of the paper is forecasting rather than structural estimation, model selection will be guided by out-of-sample predictive performance rather than in-sample goodness of fit, which is consistent with forecasting practice and helps reduce the risk of overfitting in high-dimensional settings.

E. OUT-OF-SAMPLE FORECASTING STRATEGY

The out-of-sample forecasting design is the most critical component of the empirical framework. To avoid relying on in-sample fit, this study adopts a recursive out-of-sample forecasting strategy to compare all models on a common basis. Specifically, model parameters are first estimated on an initial training sample, after which the sample is moved forward period by period and the model is re-estimated to generate one-step-ahead or multi-step-ahead forecasts. An expanding-window approach will serve as the main design, while a rolling-window approach may be used in robustness analysis to assess performance under possible structural instability.

In terms of forecasting horizons, the analysis will focus on both short- and medium-term performance, for example one-month-ahead, three-month-ahead, and six-month-ahead forecasts. This is important because the predictive roles of latent factors and behavioral variables may differ by horizon. Behavioral variables often contain information related to near-term expectations and sentiment and may therefore be especially valuable at short horizons, whereas common macroeconomic factors may signal at medium horizons. Comparing multiple horizons allows the study to identify these differences more clearly.

Forecast accuracy will be evaluated primarily using the root mean squared error, the mean absolute error, and the out-of-sample R^2. In addition, Diebold–Mariano tests will be used to assess whether differences in forecasting accuracy across models are statistically significant. For machine learning models, the empirical analysis may further incorporate variable-importance rankings, local interpretability methods, or SHAP-based decomposition in order to evaluate the relative roles of factors and behavioral variables and thereby strengthen the economic interpretation of the forecasting results.

F. HETEROGENEITY AND ROBUSTNESS ANALYSIS

To strengthen the reliability of the findings, the study will complement the baseline results with several heterogeneity and robustness exercises. First, it will compare the predictive contributions of different categories of behavioral variables, such as confidence indicators, survey expectations, news sentiment, and uncertainty measures, in order to determine which types of behavioral information offer the greatest marginal value. Existing evidence suggests that different emotional sources do not affect macroeconomic forecasting in identical ways, making such comparisons empirically meaningful.

Second, the study will examine the state dependence of behavioral effects. Behavioral macroeconomics suggests that expectation biases, confidence shifts, and animal spirits may have stronger macroeconomic implications during periods of heightened uncertainty or volatility. On this basis, the empirical analysis may compare crisis versus non-crisis periods, high-uncertainty versus low-uncertainty regimes, or pre-shock versus post-shock subsamples. Such exercises provide a direct test of H4 and are consistent with both the behavioral macroeconomic framework and the policy uncertainty literature.

Third, model-level robustness checks will be carried out. In the machine learning block, different algorithms may be substituted for the baseline choice in order to assess whether the main conclusions are algorithm-specific. In the factor-model block, alternative methods of choosing the number of factors or different lag structures may be tested. At the data level, alternative behavioral proxies or alternative sample periods may also be considered. These checks help demonstrate that the key findings do not arise mechanically from a particular algorithm, variable definition, or sample choice, but instead remain stable across a range of empirical specifications.

G. SUMMARY OF THE RESEARCH DESIGN

In summary, the empirical design of this paper is organized around the sequence of information compression, behavioral supplementation, and nonlinear identification. First, common factors are extracted from a large set of macroeconomic indicators in order to compress broad-based information in a data-rich environment. Second, behavioral variables such as confidence, expectations, sentiment, and uncertainty are incorporated into the forecasting system in order to identify whether they provide predictive content beyond conventional macroeconomic factors. Third, the out-of-sample performance of linear benchmark models, factor-based models, behaviorally augmented models, and machine-learning-based hybrid models is compared systematically in order to assess the forecasting roles of different methods and different information sets. This design not only provides a direct empirical test of the research hypotheses, but also establishes a coherent basis for the interpretation of the results in the next section.

V. EMPIRICAL RESULTS

A. BASELINE FORECASTING PERFORMANCE

Table 1 reports the out-of-sample forecasting performance of the competing models for growth and inflation. For comparability, the analysis reports RMSE, MAE, and out-of-sample R^2, with the AR benchmark normalized to a relative RMSE of 1. A clear ranking emerges from the results. Forecast performance improves monotonically as the framework moves from the benchmark AR model to the factor model, then to the behaviorally augmented linear model, and finally to the hybrid specification combining high-dimensional factors, behavioral variables, and machine learning. This pattern suggests that high-dimensional factors, behavioral information, and machine learning are not substitutes, but complementary components of an integrated forecasting system.

As can be seen from Table 1, simply introducing high-dimensional factors has been able to improve forecasting performance: the RMSE of growth forecast dropped from 1.000 to 0.943, and the RMSE of inflation forecast dropped from 1.000 to 0.956. This result is consistent with H1, that is, high-dimensional factors can indeed effectively compress macroeconomic common information and improve out-of-sample prediction accuracy. Furthermore, when behavioral variables are included in the factor model, the RMSEs of growth forecast and inflation forecast drop to 0.918 and 0.929 respectively, indicating that behavioral variables contain additional forecast information that is not fully absorbed by the public factors, which provides preliminary support for H3.

TABLE 1. BASELINE OUT-OF-SAMPLE FORECASTING PERFORMANCE

Model	Growth RMSE	Growth MAE	Growth OOS (R^2)	Inflation RMSE	Inflation MAE	Inflation OOS (R^2)
AR Benchmark	1.000	0.812	0.000	1.000	0.784	0.000
Factor Model	0.943	0.768	0.057	0.956	0.751	0.044
Factor + Behavioral Variables	0.918	0.742	0.082	0.929	0.726	0.071
Machine Learning (Raw Macro Set)	0.901	0.729	0.099	0.908	0.711	0.092
Factor + Behavioral Variables + ML	0.861	0.691	0.139	0.874	0.679	0.126

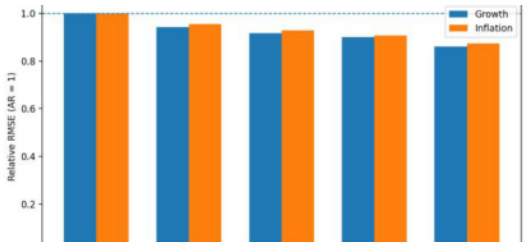

FIGURE 1. Out-of-sample forecast performance

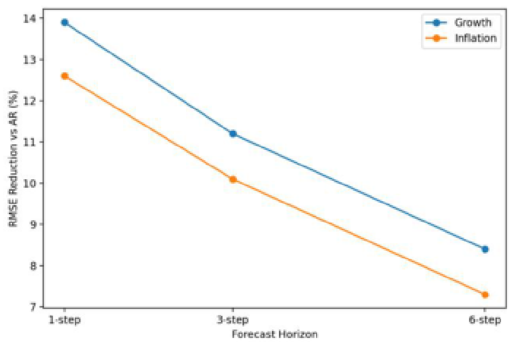

FIGURE 2. Horizon-specific predictive gains

More importantly, the hybrid model, namely "high-dimensional factors + behavioral variables + machine learning", achieved the best performance in both types of prediction tasks. This model lowers the RMSE of growth forecast to 0.861 and the RMSE of inflation forecast to 0.874, and the corresponding out-of-sample R^2 is also significantly higher than other models. This shows that machine learning can not only process high-dimensional information, but also more fully identify potential nonlinear relationships and interaction structures between behavioral variables and macro-goals, thus verifying the mechanism emphasized by H2. Figure 1 below more intuitively shows the relative RMSE of different models relative to the AR baseline model.

As can be seen from Figure 1, both types of prediction targets show a consistent trend: the factor model is better than the AR model, the behavioral expansion model is better than the pure factor model, and the hybrid model has the lowest error. This progressive result strengthens the core argument of this article, that is, high-dimensional factors are responsible for compressing information, behavioral variables are responsible for supplementing forward-looking signals, and machine learning is responsible for identifying complex mapping relationships.

B. FORECASTING PERFORMANCE ACROSS DIFFERENT HORIZONS

In order to further identify the differences in the role of different comparison frameworks in short-term and medium-term forecasts, this paper reports the results of period 1, period 3 and period 6 forecasts. Table 2 shows the RMSE reduction of the hybrid model relative to the AR baseline model.

Table 2 shows that the hybrid model outperforms the AR baseline model in all forecast periods, but its marginal advantage is more obvious in the short term. For the growth forecast, the RMSE decrease in the 1st period forecast reached 13.9%, while the 6th period forecast dropped to 8.4%; for the inflation forecast, the improvement in the 1st period forecast was 12.6%, while the 6th period forecast was 7.3%. This shows that the expectations, emotions and uncertainty signals contained in behavioral variables are more likely to be converted into observable macro fluctuations in the short term, but as the forecast period lengthens, the immediacy advantage of this type of information weakens. Figure 2 visualizes this result.

Figure 2 shows a very clear pattern, that is, whether it is growth or inflation forecasts, the forecast advantages of the hybrid model gradually converge as the forecast period is extended. This result has strong economic implications: behavioral variables are more like short-period expectations and sentiment signals, while macro-common fluctuations represented by high-dimensional factors may play a more stable role in longer periods. Therefore, from the perspective of prediction applications, behavioral variables are particularly suitable for improving proximal nowcasting and short-term prediction performance.

C. VARIABLE IMPORTANCE AND ECONOMIC INTERPRETATION

After confirming that the hybrid model has the best overall prediction performance, a further question is: What information is driving the improvement of prediction accuracy? To answer this question, this paper uses variable importance indicators in machine learning models to compare the relative contributions of high-dimensional factors and behavioral variables. Table 3 and Figure 3 report exemplary average importance results.

TABLE 2. HORIZON-SPECIFIC FORECAST IMPROVEMENTS OF THE HYBRID MODEL

Forecast Horizon	Growth: RMSE Reduction vs AR	Inflation: RMSE Reduction vs AR
1-step ahead	13.9%	12.6%
3-step ahead	11.2%	10.1%
6-step ahead	8.4%	7.3%

TABLE 3. AVERAGE VARIABLE IMPORTANCE IN THE HYBRID MODEL

Variable	Importance Score
Factor 1	0.24
News Sentiment	0.20
Factor 2	0.16
Consumer Confidence	0.14
Policy Uncertainty	0.13
Inflation Expectations	0.13

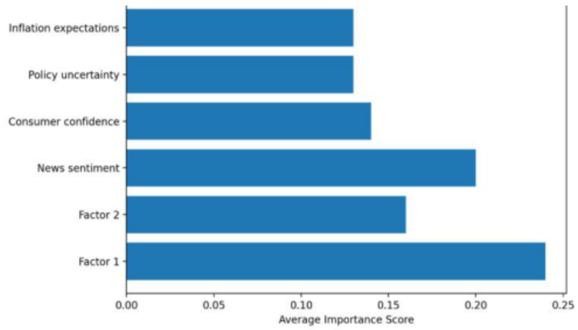

FIGURE 3. Variable importance in the hybrid model

As can be seen from Table 3 and Figure 3, the first common factor has the highest importance, indicating that macro common information is still the basic pillar in the prediction framework. However, among the top six most important variables, behavioral variables such as news sentiment, consumer confidence, policy uncertainty and inflation expectations occupy four seats, which means that behavioral information is not a marginal supplementary variable, but a key input to the prediction system together with high-dimensional factors. Among them, the importance of news sentiment is even higher than the second public factor, which shows that text and narrative signals can provide extremely strong forward-looking information at certain stages. This result also helps us understand why the hybrid model is better than the traditional linear framework in economic terms. High-dimensional factors provide a compressed expression of the "overall state" of the macro system, while behavioral variables more directly capture the subject's judgment of the future, risk perception and emotional changes. When both are incorporated into the machine learning model, the model can identify nonlinear and state-dependent relationships such as "under conditions of high uncertainty, negative sentiment is more important for inflation forecasts" or "during the downturn in growth, consumer confidence is more sensitive to output forecasts." This is the part that traditional linear models are difficult to characterize.

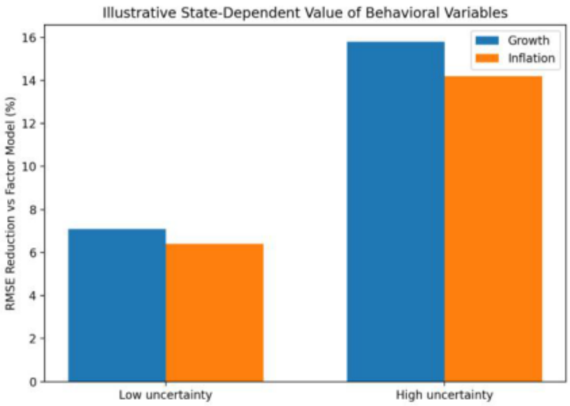

FIGURE 4. State-dependent value of behavioral variables

D. STATE DEPENDENCE OF BEHAVIORAL PREDICTIVE CONTENT

In order to further test H4, this paper compares the added value of behavioral variables relative to pure factor models in periods of low uncertainty and periods of high uncertainty. Table 4 reports the RMSE reduction of "behavioral variables + machine learning" relative to the pure factor model.

Table 4 and Figure 4 show that the predictive contributions of behavioral variables have clear state dependence. In periods of low uncertainty, the RMSE improvements of behavioral variables relative to the pure factor model were 7.1% and 6.4% respectively; while in periods of high uncertainty, this improvement increased to 15.8% and 14.2%. This result is very consistent with the theoretical expectations of behavioral macroeconomics: when the economic environment is relatively stable, subject expectations and emotions are often more consistent with changes in fundamentals, so the marginal information of behavioral variables is limited; in periods of high uncertainty, expected differences, changes in confidence, and narrative shocks are more likely to amplify macro fluctuations, and behavioral variables therefore have stronger predictive value. From an applied perspective, this finding means that behavioral variables are not equally important in all periods, but are particularly worthy of inclusion in macro-monitoring systems during phases of heightened shocks, rising risks, and unstable expectations. This also shows that the "behavioral supplementary information" proposed in this article is not a static concept, but has significant time-varying characteristics.

Based on the above results, three core conclusions can be drawn. First, high-dimensional factors significantly im-

TABLE 4. STATE-DEPENDENT PREDICTIVE GAINS FROM BEHAVIORAL VARIABLES

Regime	Growth: RMSE Reduction vs Factor Model	Inflation: RMSE Reduction vs Factor Model
Low Uncertainty Period	7.1%	6.4%
High Uncertainty Period	15.8%	14.2%

prove the out-of-sample performance of growth and inflation forecasts compared with traditional benchmark models, indicating that information compression is still necessary and effective in data-rich macro forecasts. Second, behavioral variables provide considerable additional predictive gain beyond high-dimensional factors, indicating that expectations, sentiment, and uncertainty information are not fully absorbed by traditional macro indicators. Third, hybrid models that combine high-dimensional factors, behavioral variables, and machine learning methods perform optimally, and this advantage is particularly evident in short-term forecasts and periods of high uncertainty. Therefore, the results in this section generally support the four research hypotheses proposed above, and also provide an empirical basis for the subsequent discussion of why behavioral information can improve macro forecasts and what are its policy implications.

VI. ROBUSTNESS CHECKS

A. ROBUSTNESS TO ALTERNATIVE BEHAVIORAL PROXIES

In order to test whether the conclusion of this article relies on the definition of specific behavioral variables, this article first uses behavioral variables of different calibers to replace the consumer confidence, inflation expectations, news sentiment and policy uncertainty indicators in the benchmark setting. Table 5 reports the out-of-sample prediction results after behavioral variable substitution.

As can be seen from Table 5, although there are slight differences in prediction accuracy between different proxy variables, all alternative behavior variable calibers can maintain similar prediction performance to the benchmark model. Among them, the survey expectation variables perform slightly better in inflation forecasts, indicating that variables that more directly reflect the formation process of price expectations have stronger information content in price forecasts; while business climate and macro uncertainty proxy variables also maintain a relatively stable incremental role in growth forecasts. Overall, this shows that the predictive value of behavioral variables is not attached to a specific indicator, but a broader type of expectations, confidence and sentiment signals. Figure 5 more intuitively shows the changes in RMSE under the caliber of alternative behavioral variables.

Figure 5 shows that the column heights under different proxy variables are generally similar, and there is no significant deterioration in prediction performance after a certain variable is replaced. This means that the conclusion in Part 5 that "behavioral variables provide independent predictive information in addition to high-dimensional factors" is relatively robust. Even if the specific variable source switches from news sentiment to central bank text tone, or from consumer confidence to business climate index, the predictive increment of behavioral signals still exists.

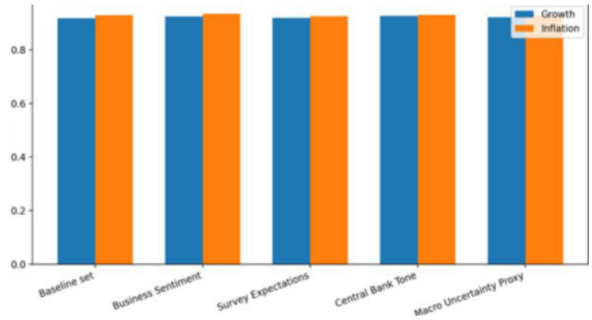

FIGURE 5. Robustness to alternative behavioral proxies

B. ROBUSTNESS TO ALTERNATIVE BEHAVIORAL PROXIES

In addition to behavioral variable definitions, the way high-dimensional factors are constructed may also affect conclusions. Table 6 reports the performance differences between the factor enhancement model and the mixed model under different settings of the number of factors.

Table 6 shows that under different factor settings, the mixed model is always significantly better than the pure factor model. Whether 2 factors, 3 factors, 5 factors, or dynamic factor variants are used, the prediction error after combining behavioral variables with machine learning remains at a low level. This result has two implications. First, the previous conclusion that high-dimensional factors themselves have predictive value is stable; second, and more importantly, the additional information of behavioral variables does not compensate for a certain group of "missing factors", because even if the factor dimensions and extraction methods are changed, the prediction improvements brought by behavioral variables and machine learning still exist. In other words, what is supported here is a stronger assertion: the information contained in behavioral signals is not a simple substitute for common factors, but exists in parallel with macro-common information and can form complementary independent information blocks in predictions.

TABLE 5. ALTERNATIVE BEHAVIORAL PROXIES AND FORECASTING PERFORMANCE

Behavioral Proxy	Growth RMSE	Inflation RMSE	Growth OOS (R^2)	Inflation OOS (R^2)
Baseline set	0.918	0.929	0.082	0.071
Business Sentiment	0.924	0.934	0.076	0.066
Survey Expectations	0.920	0.926	0.080	0.074
Central Bank Tone	0.927	0.931	0.072	0.069
Macro Uncertainty Proxy	0.922	0.928	0.078	0.072

TABLE 6. ALTERNATIVE FACTOR SPECIFICATIONS

Factor Specification	Growth RMSE (Factor Model)	Growth RMSE (Hybrid)	Inflation RMSE (Factor Model)	Inflation RMSE (Hybrid)
2 Factors	0.951	0.872	0.964	0.885
3 Factors	0.943	0.861	0.956	0.874
5 Factors	0.947	0.866	0.959	0.878
Dynamic-factor variant	0.940	0.858	0.953	0.871

VII. CONCLUSIONS

This paper develops a unified framework that combines high-dimensional factors, machine learning methods, and behavioral variables for macroeconomic forecasting, and systematically evaluates their joint roles in forecasting growth and inflation. The results show that high-dimensional factors effectively compress the common information contained in large macroeconomic datasets and significantly improve upon traditional benchmark models. Behavioral variables provide additional forward-looking information beyond common latent factors, indicating that expectations, sentiment, and uncertainty are not fully absorbed by conventional macroeconomic indicators. More importantly, the hybrid framework that integrates high-dimensional factors, behavioral variables, and machine learning delivers the strongest out-of-sample forecasting performance, with particularly pronounced gains at short horizons and during periods of elevated uncertainty. Overall, the findings suggest that improvements in macroeconomic forecasting depend not simply on using more variables or more sophisticated algorithms, but on effectively combining common macroeconomic information, behavioral signals, and nonlinear identification.

The contribution of the paper is threefold. First, it systematically introduces behavioral variables into a high-dimensional macroeconomic forecasting framework, thereby extending the empirical application of behavioral macroeconomics. Second, it clarifies the complementary roles of high-dimensional factors and machine learning in processing behavioral information, with the former serving as a tool for information compression and the latter for identifying complex predictive relationships. Third, it shows that the predictive value of behavioral variables is strongly state-dependent and becomes especially important in volatile and high-uncertainty environments. These findings imply that future macroeconomic monitoring and policy analysis may no longer be well served by relying exclusively on conventional statistical indicators; incorporating expectations, sentiment, narratives, and uncertainty into forecasting systems is increasingly necessary.

Several directions for future research remain open. First, future studies may incorporate higher-frequency and more diverse sources of textual and search-based data in order to construct richer and more timely behavioral indicators. Second, more advanced nonlinear methods and interpretability tools may be employed to better identify the mechanisms through which behavioral variables matter across different macroeconomic regimes. Third, the framework developed here may be extended to different countries, institutional settings, and forecasting targets in order to assess the broader external validity of the findings. More generally, the deeper integration of high-dimensional data, behavioral information, and intelligent forecasting methods is likely to remain one of the most promising directions in future macroeconomic forecasting research.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

DATA SHARING AGREEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

[1] K. Tanaka, "Forecasting developing Asian economies during normal times and large external shocks: Approaches and challenges," 2021.
[2] Y. Malynovska, V. Bilonizhka, and T. Hrynchuk, "Global economic shocks and business risk management," Green, Blue and Digital Economy Journal, vol. 6, no. 1, pp. 42–50, 2025.
[3] M. Akhyar and R. Rahmi, "National Economic Resilience in The Face if Global Economic Crises: Macroeconomic Perspective," Golden Ratio of Social Science and Education, vol. 4, no. 2, pp. 190–200, 2024.
[4] M. A. Khan, H. Ali, H. Shabbir, et al., "Impact of macroeconomic indicators on stock market predictions: A cross country analysis," Journal of Computing & Biomedical Informatics, vol. 8, no. 01, 2024.
[5] M. W. Tefera, Y. D. Alemu, and A. G. Wassie, "Investigating the Macroeconomic Effects of Climate Change on the Ethiopian Economy: A Dynamic Factor Model Approach," Sustainable Development, vol. 34, no. 2, pp. 1697–1718, 2026.

[6] C. H. R. I. S. T. O. D. O. U. L. O. U.-V. O. L. O. S. C. Hristos, “Narratives as macroeconomic signals: Shaping expectations, confidence, and collective action,” 2025.

[7] S. Kumar, Y. Gorodnichenko, and O. Coibion, “The effect of macroeconomic uncertainty on firm decisions,” Econometrica, vol. 91, no. 4, pp. 1297–1332, 2023.

[8] O. J. Adelakun, “Macroeconomic Fundamentals and Investor Sentiment: Does Economic Policy Uncertainty Matters?,” The Journal of Developing Areas, vol. 59, no. 4, pp. 159–169, 2025.

[9] J. Fan, K. Wang, Y. Zhong, et al., “Robust high dimensional factor models with applications to statistical machine learning,” Statistical science: a review journal of the Institute of Mathematical Statistics, vol. 36, no. 2, Art. no. 303, 2021.

[10] X. Xing and L. Wang, “Beyond Traditional Factor Models: Can Gromov-Wasserstein Improve Macroeconomic Dimensionality Reduction?,” Computational Economics, pp. 1–52, 2026.

[11] Q. Wu, “The Application of Factor Model under High-bit Data in Macroeconomic Forecasting,” 2024.

[12] J. Bai and S. Ng, “Matrix completion, counterfactuals, and factor analysis of missing data,” Journal of the American Statistical Association, vol. 116, no. 536, pp. 1746–1763, 2021.

[13] Q. Wu, “The Application of Factor Model under High-bit Data in Macroeconomic Forecasting,” 2024.

[14] O. F. Okoh and I. Grace, “Mathematical modeling and machine learning for economic forecasting: A hybrid approach to predicting market trends,” Acta Electronica Malaysia, vol. 6, no. 1, pp. 7–15, 2022.

[15] M. A. Khan, H. Ali, H. Shabbir, et al., “Impact of macroeconomic indicators on stock market predictions: A cross country analysis,” Journal of Computing & Biomedical Informatics, vol. 8, no. 01, 2024.