5. CAN AI-DRIVEN NATIONAL ESG IN BIG DATA HELP IMPROVE MACROECONOMIC FORECASTING?[1]

Hao XIAO[2]
Xiaofen LI[3]
Junyi YANG[4]
Xinjian YE[5*]

Abstract

This study introduces a novel predictive framework for predicting South Africa's macroeconomic trends using national ESG in big data based on AI technology and deep learning. This study utilizes the GDELT database and AI-driven indicator construction methods to extract meaningful insights from 10.76 million news, generating ESG in big data at the national governance level. By combining traditional macroeconomic indicators with national ESG in big data, this study evaluates the predictive performance of econometric, machine learning, and deep learning models. The rolling out-of-sample prediction analysis shows that the LSTM model achieves the highest prediction accuracy. Subsequently, LSTM models with and without national ESG in big data were designed to evaluate the extent to which incorporating national ESG in big data improves prediction accuracy. This study demonstrates that national ESG in big data enhances the accuracy of macroeconomic forecasting, particularly improving the short-term forecasting performance of the models.

Keywords: *macroeconomic forecasting, national ESG index, news data, AI, LSTM*

JEL Classification: *C53, E17, Q56, O33*

---

[1] *Paper from the special issue "Artificial Intelligence for Enhanced ESG Integration in Economic Forecasting", Guest Editors: Chi-Wei SU and Malin SONG.*

* *This research is supported by National Natural Science Foundation of China (72341021) and Hunan Provincial Natural Science Foundation of China (2023JJ30172). The authors have contributed equally to this work.*

[2] *School of Economics & Trade, Hunan University; Institute of African Studies, Hunan University, Changsha, Hunan Province, China, 410079. Email: xh_26@126.com*

[3] *School of Economics & Trade, Hunan University, Changsha, Hunan Province, China, 410079, Email: xflixian@163.com*

[4] *School of Economics & Trade, Hunan University, Changsha, Hunan Province, China, 410079, Email: youngbayia@hnu.edu.cn*

[5] *School of Economics & Trade, Hunan University, Changsha, Hunan Province, China, 410079, Email: yexinjian0203@126.com, *corresponding author*

1. Introduction

In the 30 years since the establishment of the new South Africa, the country's gross domestic product (GDP) has generally trended upward. From the end of apartheid in 1994 to 2010, South Africa's annual economic growth rate averaged around 3-5%. South Africa has abundant mineral resources, especially in gold and platinum mining, which has made a significant contribution to economic development. Meanwhile, the diversification of the manufacturing and the expansion of services, particularly in finance, tourism, and information technology, have further promoted economic growth, gradually making South Africa one of Africa's most developed economies. However, South Africa's economic development still faces numerous risks and challenges. In recent years, South Africa's economic growth rate has slowed significantly, from 2015 to 2019, the annual growth rate remained around 1%, below the global average, limiting the government's ability to invest in social welfare and infrastructure. In addition, the South African economy is constrained by multiple internal factors, including high unemployment, energy shortages, and structural inefficiencies within the economy, which have weakened overall consumer demand, thereby undermining economic dynamism and social stability. On the other hand, as a representative open economy in Africa, South Africa faces significant economic uncertainty due to various external factors, including rising commodity prices and geopolitical risks. Accurate and timely macroeconomic forecasting can help the South African government and businesses to formulate strategies, allocate resources, and implement responsive measures more effectively in an uncertain global environment.

There are a large number of scholars working on how to develop more effective macroeconomic forecasting methods. The primary approaches to macroeconomic forecasting include econometric models (Hansson et al., 2005; Gupta and Kabundi, 2011; Balcilar et al., 2015; Yemba et al., 2023), artificial intelligence algorithms (Siami et al., 2018; Gu et al., 2020; Han et al., 2023; Lupu et al., 2024), and integrated models (Wei et al., 2010; Barrow and Crone, 2016; Godahewa et al., 2023). In recent years, with the widespread adoption of internet technology, micro-level dynamics such as news and platform economies have emerged as significant drivers of macroeconomic changes (Gentzkow et al., 2019). Advanced technologies, including natural language processing (NLP) and image recognition, enable the extraction of information from unstructured textual data, such as news and social media, providing high-frequency signals that offer valuable insights for national economic analysis.

Big data technology is revolutionizing the traditional paradigm of economic forecasting, propelling it to a new phase of development (Lin and Wei, 2024). Compared to traditional economic statistical indicators, big data encompasses a broader range of metrics, including real-time data and extensive unstructured data, thereby offering a more diverse and enriched data source for macroeconomic forecasting. Unstructured textual data on public opinion provides dynamic and comprehensive information, offering substantial data support for macroeconomic analysis. The wealth of information embedded in these data sources plays a critical role in macroeconomic forecasting (Kohns and Bhattacharjee, 2023; Ashwin et al., 2024). A review of the existing literature indicates that most studies on forecasting macroeconomic trends using public opinion data predominantly rely on news as the primary source of such information. For instance, Larsen and Thorsrud (2019) analyzed the ability of news to explain and predict macroeconomic fluctuations by extracting topics and sentiments from news content and applying the SVAR model. Kalamara et al. (2022) incorporated economic signals derived from news texts into forecasting models to predict key macroeconomic variables, including GDP, inflation, and unemployment. Thorsrud (2020) and Zheng et al. (2024) integrated traditional macroeconomic indicators with news data to predict macroeconomic trends, highlighting the distinct advantages of textual data in economic forecasting. Hong et al. (2024) employed a large-scale news corpus and machine learning algorithms to analyze news content and extract topics, which were then used to forecast inflation. A limited number of studies have utilized public opinion data sources other than news for macroeconomic forecasting. For instance, Bratu and Nicula (2024) transformed extensive textual data from inflation reports into sentiment indices to aid in macroeconomic predictions. Similarly, Lin et al. (2023) integrated central bank communication texts with macroeconomic forecasting models.

As evidenced by the aforementioned literature, the majority of studies rely on news texts as public opinion data for macroeconomic forecasting. In response, this paper introduces an innovative approach by leveraging big data in the macroeconomic forecasting domain, proposing a high-frequency national ESG indicator that integrates artificial intelligence techniques with textual big data. Additionally, it examines how enhancing the LSTM model can improve the precision and practical utility of macroeconomic forecasting.

The paper proceeds as follows. Section 2 presents the development and processing of AI-driven national ESG in big data. Section 3 describes the design and evaluation of the prediction models. Section 4 presents the empirical analysis. Section 5 summarizes the research findings.

2. AI-Driven National ESG in Big Data

2.1. Problem Statement

The World Bank has proposed a framework for developing national-level ESG indicators, which encompasses three key dimensions: environmental, social, and governance. In the environmental dimension, national ESG focuses on the externalities of economic activities, including sustainable resource access, pollution control, and climate response. These indicators aid in assessing a country's sustainability in environmental protection and resource management. In the social dimension, national ESG emphasizes meeting basic population needs, ensuring equity, and reducing poverty—factors essential to supporting a country's economic growth. For governance, national ESG measures social stability and government effectiveness in governance. However, accurately measuring national ESG remains challenging due to limitations in data availability and the lack of informative high-frequency indicators. Although the World Bank's indicator system provides data support for constructing a national ESG index, much of this data is available only on an annual basis, with few monthly or higher-frequency sources. This limits the application of national ESG indices for short-term trend analysis and real-time monitoring (Capelle et al., 2019; Jiang et al., 2022). In the rapidly changing international political and economic environment, the limitation in data frequency constrains the ability of national ESG indices to respond to short-term fluctuations. To address these limitations, this paper introduces high-frequency public sentiment data and combines it with AI-driven dynamic analysis for national ESG monitoring. Public sentiment data can rapidly capture the levels of public and media attention to various events and shifts in sentiment, providing real-time insights into public opinion trends and valuable feedback for governance. As an early warning indicator in national governance, sentiment data can reveal potential social risks, enabling governments and policymakers to detect changes in public sentiment promptly and respond quickly to prevent and mitigate potential sources of social instability (Voukelatou et al., 2022; Su et al., 2024). Therefore, this study uses public sentiment data as a proxy indicator for national governance, constructing a real-time, dynamic national ESG index by analyzing public sentiment from sources such as news events and social media.

2.2. The National ESG and the Macroeconomy

National ESG is closely related to macroeconomic performance, as it reflects a country's sustainability, social stability, and governance standards, all of which significantly influence both long-term economic growth and short-term macroeconomic volatility (Naomi and Akbar, 2021; Qin et al., 2024).

The economic impact of national ESG initiatives is primarily observed in areas such as trade and taxation. Environmental, social, and governance policies at the national level can strengthen a country's competitiveness in attracting foreign direct investment (Zhang et al,2022;Parikh et al,2023). By prioritizing national ESG standards, governments aim to establish a stable and transparent regulatory environment that mitigates investment risks and bolsters investor confidence. This influx of investments, in turn, fosters economic growth and serves as a driving force for the country's broader economic development. The generation of tax revenue and government income relies on the presence of a stable political and economic environment (Umar et al., 2020; Hübel, 2022). Countries demonstrating strong ESG performance mitigate ESG-related risks at the national level by prudently managing natural resources, promoting equitable and sustainable development, enhancing human capital, and establishing robust political and financial standards. These policies contribute to stabilizing tax revenues and reducing the risks associated with adverse events such as natural disasters, political instability, or corruption scandals.

2.3. Construction of the National ESG in Big Data

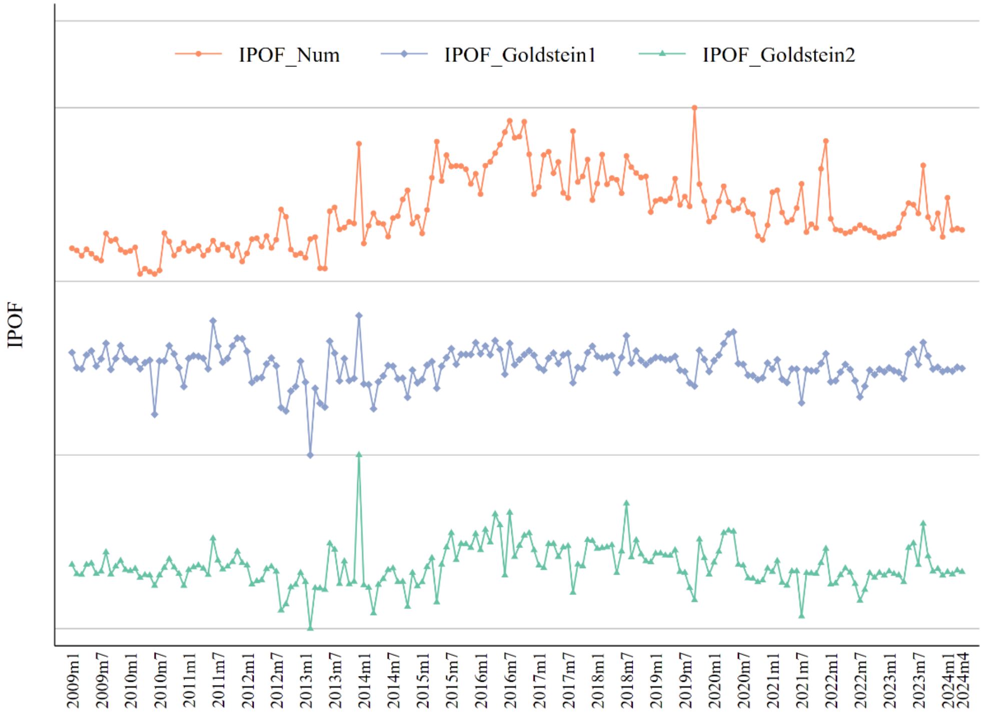
The GDELT (Global Database of Events, Language, and Tone) database[6] is a large-scale, global dataset designed to monitor, record, and analyze events, news, and sentiment information on a worldwide scale. GDELT continuously collects textual data from various sources, including news reports, social media, radio, and television broadgatecasts, from around the globe. It processes billions of news articles each year, covering over 65 languages across a wide range of domains such as politics, economics, society, and culture. Utilizing advanced automated techniques, it extracts metadata from global broadcast, print, and online news sources and translates this information into English in near real-time. The database is updated every 15 minutes on a dedicated online platform (Coniglio et al., 2023; Sun et al., 2021). Based on the Conflict and Mediation Event Observation (CAMEO) event coding system, each event is associated with two actors (Actor1 and Actor2), and each event is assigned a set of attributes referred to as the Goldstein Scale, which ranges from -10 to 10. This scale is used to quantify the degree of conflict or cooperation between the two actors involved in the event (Consoli et al., 2021; Elshendy et al., 2018; Shen et al., 2022). This paper systematically analyzes daily records of interaction events involving South Africa and other countries, extracted from the GDELT 2.0 database for the period from January 2009 to May 2024. Each event includes comprehensive metadata, providing detailed information on time, location, participants, and event type.

As a real-time indicator within national ESG in big data, IPOF (International Public Opinion Factor) indicators help address the challenges of obtaining traditional ESG data and its low update frequency, making them particularly suitable for high-frequency data needs. This paper employs three distinct methods to construct a national ESG big data index to measure South Africa’s level of governance. The first method is based on the NumMentions variable, which indicates the frequency with which an event is mentioned across all articles in the database, thus serving as an indicator of the event’s importance. In line with the approach of Baker et al. (2016), this method uses the monthly count of relevant mentions to construct the national ESG in big data index (IPOF_Num). The second and third methods are inspired by Pascal & Georg (2015). The Goldstein Scale assigns each event an impact score within a range of [-10, 10], with positive values indicating cooperative events and negative values indicating conflict-related events. By calculating the monthly average and total Goldstein scores, two national ESG big data indices are constructed (IPOF_Goldstein1 and IPOF_Goldstein2). Furthermore, the QuadClass variable can be used to distinguish event types, enabling the construction of additional international

[6] *The GDELT dataset can be downloaded from https://www.gdeltproject.org/ (accessed on January 23, 2025).*

sentiment factors by event category. The trend of South Africa Monthly National ESG in Big is illustrated in Figure 1.

Figure 1. Monthly of South Africa National ESG in Big Data Index



3. Model Specification

This section provides an overview of the models used for macroeconomic forecasting, including econometric models, machine learning models, and deep learning models. To compare the performance of these models, we use Root Mean Square Error (RMSE) as the primary evaluation metric.

3.1. Econometric Forecasting Model

3.1.1 VAR

The Vector Autoregression (VAR) model, proposed by Sims (1980), is capable of capturing the dynamic relationships among multiple variables. The general expression for the VAR model .

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + c + \mu_t, t = 1, 2, \cdots, T. \quad (1)$$

In this equation, y_t denotes a k dimensional column vector of endogenous variables, including macroeconomic indicators such as GDP, CPI, and M2. A_1 to A_p are $m \times m$ coefficient matrices, where p is the lag order, c is an m dimensional vector, μ_t is a k dimensional vector of disturbance terms. Expanding equation (1) results in:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{pmatrix} = A_1 \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{kt-1} \end{pmatrix} + A_2 \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \\ \vdots \\ y_{kt-k} \end{pmatrix} + \cdots + A_p \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ \vdots \\ y_{kt-p} \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_t \end{pmatrix} + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{kt} \end{pmatrix}, t = 1, 2, \cdots, T. \quad (2)$$

3.1.2 FAVAR

The Factor-Augmented Vector Autoregression (FAVAR) model is an extension of the Vector Autoregression model. By applying principal component analysis (PCA) or factor analysis, the FAVAR model extracts latent factors to capture information from high-dimensional data more comprehensively, improving the analysis of dynamic relationships among multiple variables (Bernanke et al., 2005). The FAVAR model is as follows:

$$\tilde{y}_t = B_1 \tilde{y}_{t-1} + B_2 \tilde{y}_{t-2} + \cdots + B_p \tilde{y}_{t-p} + c + \mathrm{v}_t, t = 1, 2, \cdots T. \quad (3)$$

$$\begin{pmatrix} y_t \\ F_t \end{pmatrix} = A_1 \begin{pmatrix} y_{t-1} \\ F_{t-1} \end{pmatrix} + A_2 \begin{pmatrix} y_{t-2} \\ F_{t-2} \end{pmatrix} + \cdots + A_p \begin{pmatrix} y_{t-p} \\ F_{t-p} \end{pmatrix} + c + v_t, t = 1,2,\cdots,T \quad (4)$$

In this equation, $\tilde{y}_t = \begin{pmatrix} y_t \\ F_t \end{pmatrix}$ is a combined vector, where y_t represents a k dimensional vector of endogenous macroeconomic variables, including indicators such as GDP, CPI, and M2. F_t denotes a q dimensional vector of latent factors that capture additional unobserved information, $B_1, B_2, \cdots, B_p$ are coefficient matrices, c is an mmm-dimensional vector, v_t is the combined noise term.

3.2.Machine Learning Prediction Model

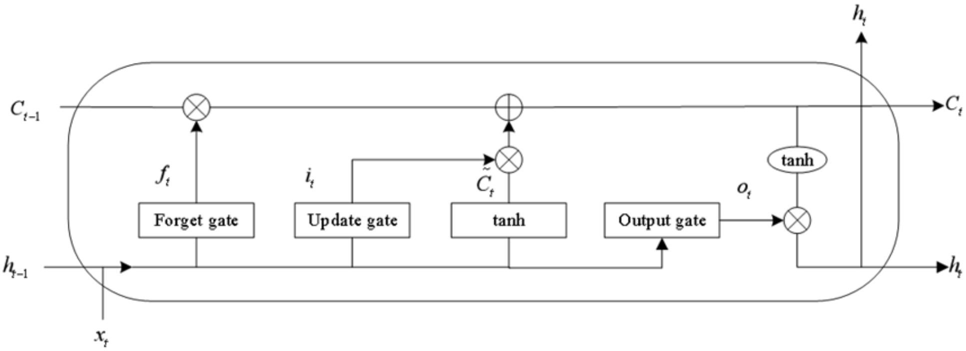
Support Vector Regression (SVR) is a regression technique based on the principles of Support Vector Machines (SVM) and is applied to predict continuous target variables. In regression tasks, SVR constructs a model that minimizes the error for most training samples relative to a prediction function, ensuring that this error remains within a preset threshold. Concurrently, SVR seeks to maximize the margin between support vectors in the feature space.

The core concept of SVR is to identify an optimal function that minimizes the error for most training samples within a preset threshold, while simultaneously maximizing the margin between support vectors. This approach is effective in modeling nonlinear relationships by selecting an appropriate kernel function to map input data into a high-dimensional feature space, thereby facilitating the capture of complex patterns.

3.3. Deep Learning Prediction Model

The Long Short-Term Memory (LSTM) network is a widely used deep learning prediction model, well-suited for capturing long-term dependencies and frequently applied to a variety of complex problems (Hochreiter and Schmidhuber, 1996). The distinct feature of the LSTM network lies in its use of memory cells and nonlinear gating units. Memory cells allow the model to retain historical states over time, while nonlinear gating units control the input and output of information, effectively addressing issues related to long-term dependencies. As illustrated in Figure 2, the LSTM model processes information through the coordinated operation of three gates (the forget gate, input gate, and output gate) to perform time series predictions, denoted as h_t.

Figure 2. Structure of an LSTM Cell



To implement the signal forgetting process, the forget gate uses a Sigmoid activation layer to determine the extent to which the signal is retained. The forget gate receives the current input signal x_t and the output signa h_{t-1} from the previous layer. Thus, the output of the forget gate at time t can be expressed as:

$$f_t = \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}\right) \tag{5}$$

The information stored in the memory cell of an LSTM network comprises two components. The first component, i_t is the result produced by the Sigmoid layer of the input gate, representing new information that needs to be updated. The second component, u_t is a candidate vector generated by the tanh layer, which is added to the cell state. Building on the previous memory cell state C_{t-1} , this structure facilitates the memory update process in the LSTM. The cell state update involves an element-wise product between $i_t \otimes u_t$, resulting in the new memory cell state.

$$i_t = \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}\right) \tag{6}$$

$$u_t = \tanh\left(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}\right) \tag{7}$$

$$C_t = i_t \otimes u_t + f_t \otimes C_{t-1} \tag{8}$$

The final output of the LSTM model is determined by the output gate. The Sigmoid layer modulates the output from the memory cell, while the tanh function processes the signal within the cell state. These two parts are then combined through element-wise multiplication to produce the final output.

$$h_t = o_t \otimes \tan\left(C_t\right) \tag{9}$$

In this context, x_t denotes the input vector at time t, which includes historical data for macroeconomic variables such as GDP, CPI, and M2. h_{t-1} represents the previous output, and C_{t-1} is the memory cell state from the prior time step. The forget gate determines how much of the previous memory to discard, while the input gate decides how much new information to add. The final cell state is then formed by integrating new and retained information. Lastly, the output gate extracts the most relevant information from the final memory state for prediction.

3.4. Evaluation indicators (RMSE)

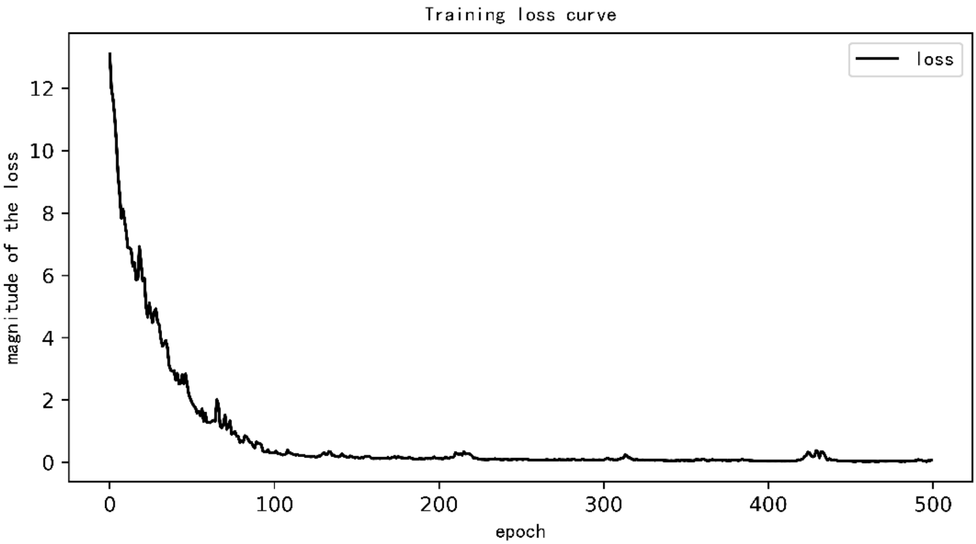
In this study, the commonly used out-of-sample Root Mean Squared Error (RMSE) from existing literature (Hansen and Timmermann,2015) is used to measure prediction ability. RMSE is a widely used statistic that reflects the accuracy of a model by measuring the difference between predicted values and actual observations. After the model is trained, out-of-sample predictions are performed using the training set. At this point, the model has not seen the data in the test set during prediction, thus reflecting the model's generalization ability more realistically. the smaller the RMSE value, the better the model's prediction effect.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{10}$$

n is the number of test sample. y_i and $\hat{y}_i$ represent the actual value of the test set and the predicted value of the model, respectively.

3.5. Optimizer based on LSTM

Figure 3. Convergence of the GDP loss function as a function of the number of iterations



In this paper, the Adam optimizer is used for model training. The advantage of the Adam optimizer is that it combines momentum and adaptive learning rate adjustment, which can correct the gradient deviation during the training process so that the learning rate of each iteration fluctuates within a reasonable range, thus improving the convergence of the parameters and the stability of the training process. Taking the South African GDP prediction as an example, this experiment set the learning rate to 0.01 and performed 100 iterations. As can be seen from the training loss curve (Training loss curve), the loss value decreases rapidly at the beginning of training, which indicates that the Adam optimizer effectively helps the model to learn the data features and gradually levels off after about 100 iterations, and the model has stabilized at 500, so the experiment sets the number of iterations to 500. Figure 3 illustrates the declining loss of the South African GDP prediction, with the parameters of the other predictor variables dynamically adjusted according to model convergence.

4. Empirical research

4.1. Data sources and processing

This study selects six macroeconomic variables as target variables for forecasting: the Consumer Price Index (CPI), Producer Price Index (PPI), year-over-year growth rate of the broad money supply (M2), year-over-year growth rate of exports (Exp), year-over-year growth rate of imports (Imp), and year-over-year GDP growth rate (GDP). GDP serves as a crucial indicator of economic performance. To align with monthly indicators, this study employs a linear interpolation method to convert quarterly data into higher-frequency monthly data, enabling the calculation of the monthly real GDP growth rate. South Africa’s macroeconomic development exhibits unique historical characteristics that differ significantly from those of developed Western countries and other African nations.

Table 1. Selection of variables and basis

<table>
<tr><th>Indicator name</th><th>variables</th><th>Data sources</th><th>Sources</th></tr>
<tr><td>Year-on-year real GDP growth</td><td>GDP</td><td rowspan="3">Statistics South Africa</td><td rowspan="6">Gupta and Steinbach(2013)；Kabundi and Mbelu(2021)</td></tr>
<tr><td>CPI year-on-year growth rate</td><td>CPI</td></tr>
<tr><td>year-on-year growth rate of the broad money supply</td><td>M2</td></tr>
<tr><td>Producer Price Index</td><td>PPI</td><td rowspan="3">South African Revenue Service</td></tr>
<tr><td>year-on-year growth rate of imports</td><td>Imp</td></tr>
<tr><td>year-on-year growth rate of exports</td><td>Exp</td></tr>
<tr><td>The real effective exchange rate</td><td>Er</td><td rowspan="3">South African Reserve Bank</td><td rowspan="6">Gupta and Steinbach(2013)；Kabundi and Mbeluet(2021)；Liu and Cheng(2024)</td></tr>
<tr><td>The Retail Trade Index</td><td>Rt</td></tr>
<tr><td>The Macro-Economic Leading Indicator</td><td>LI</td></tr>
<tr><td>The electricity generation index</td><td>ELEC</td><td>Statistics South Africa</td></tr>
<tr><td>The U.S. Industrial Production Index</td><td>US IPI</td><td rowspan="2">US Federal Reserve</td></tr>
<tr><td>The U.S. Federal Funds Rate</td><td>US r</td></tr>
<tr><td>National ESG in big data</td><td>IPOF_Num</td><td>The Global Database</td><td>Baker et al.(2016)</td></tr>
</table>

Accordingly, this paper identifies suitable forecasting indicators based on the characteristics and level of South Africa's economic development. To comprehensively capture the macroeconomic characteristics of South Africa, this study includes additional factors such as the composite leading economic index, electricity production index, and real exchange rate. The specific indicators are as follows: (1) The Leading Indicator of macroeconomic conditions, published by the South African Reserve Bank, which is a widely recognized and authoritative leading economic indicator frequently applied in macroeconomic forecasting. (2) The electricity production index, an important measure of economic health that reflects industrial production activity and factory operating levels in the region. (3) The real effective exchange rate, which measures the true purchasing power of the country's currency and is a key factor impacting the macroeconomy. (4) The retail trade index, an economic measure reflecting retail price trends for goods across urban and rural areas. In an open economy, South Africa maintains frequent trade and investment interactions with major economies such as Europe and China, making its economy directly affected by changes in the global economic environment. This study follows the indicator selection approach of Higgins et al. (2016) by including the U.S. Industrial Production Index and the U.S. Federal Funds Rate, aiming to capture South Africa's role within the global economic system. Based on data availability, the timeframe for this study's data spans from January 2009 to April 2024. Details on the variable names, data sources, and relevant information are provided in Table 1.

4.2. Comparison of Prediction Model Performance

Following a systematic analysis of commonly used prediction models, this study compares the Vector Autoregression (VAR) model, the Factor-Augmented Vector Autoregression (FAVAR) model, the Support Vector Regression (SVR) model, and the Long Short-Term Memory (LSTM) model to identify a suitable benchmark model. This study uses data from January 2009 to April 2024 as the sample period, with January 2021 to April 2024 serving as the test set for forecasting. The target variables are forecasted at 1-period, 3-period, 6-period, and 12-period horizons. To evaluate and compare the predictive performance of each model, rolling out-of-sample RMSE was calculated, as shown in Table 2. Table 2 shows that the LSTM model performs optimally or near-optimally in most cases, underscoring the unique advantages of deep learning models in processing complex data. The SVR model ranks second, indicating that, in South African macroeconomic forecasting, deep Learning and machine learning generally outperform traditional econometric models. Among the econometric models, the VAR model has the lowest prediction accuracy, primarily due to its limited sample size, which constrains its ability to integrate large-scale information, resulting in poorer performance in macroeconomic forecasting. By comparing the VAR model, FAVAR model, and SVR model, this study identifies the LSTM model as having the best performance. Consequently, the LSTM model is selected as the benchmark to investigate the contribution of national ESG in big data to predicting macroeconomic trends in South Africa.

Table 2. Performance Comparison of Various Prediction Models

target variable	Predicted number of steps τ	VAR	FAVAR	SVR	LSTM
GDP	1	3.08	2.95	2.40	2.70
	3	5.30	5.11	4.31	4.16
	6	5.42	5.2	4.03	2.93
	12	4.25	3.94	3.87	4.69
Cumulative RMSE		18.05	17.2	14.61	14.48

target variable	Predicted number of steps τ	VAR	FAVAR	SVR	LSTM
CPI	1	0.44	0.44	0.67	0.47
	3	1.02	0.98	0.79	0.76
	6	1.71	1.64	0.81	0.95
	12	2.55	2.44	0.95	0.97
Cumulative RMSE		5.72	5.5	3.23	3.15
M2	1	1.62	1.57	1.56	2.24
	3	2.80	2.56	2.59	2.98
	6	4.29	3.83	2.60	2.05
	12	5.09	4.51	3.21	2.18
Cumulative RMSE		13.80	12.47	9.96	9.45
Imp	1	11.21	12.01	14.44	8.76
	3	13.55	14.15	15.74	14.36
	6	17.50	17.40	14.51	15.12
	12	21.75	20.99	14.77	19.76
Cumulative RMSE		64.01	64.55	59.45	58.00
Exp	1	50.22	48.06	33.59	31.29
	3	61.47	57.27	33.38	32.18
	6	59.25	54.71	30.82	34.58
	12	48.69	44.42	31.96	34.42
Cumulative RMSE		219.63	204.46	129.74	132.47
PPI	1	3.76	3.75	4.85	4.49
	3	4.00	4.16	4.72	4.51
	6	4.01	4.37	5.22	4.47
	12	4.17	4.35	4.00	4.60
Cumulative RMSE		15.94	16.63	18.80	18.07

Note: This table is based on the test set from January 2021 to April 2024, calculating the rolling out-of-sample prediction errors (RMSE) for various macroeconomic forecasting models. The cumulative RMSE denotes the total of the predicted RMSEs.

4.3. Comparative Analysis of Forecasting Accuracy Based on National ESG in Big Data

Considering the exceptional predictive capability of the LSTM model, this section adopts it as the benchmark to examine how national ESG in big data enhances and optimizes the accuracy of macroeconomic forecasting. Using the rolling pseudo out-of-sample forecasting method of Kelly et al. (2019), this study initially uses a subset of the sample data as the training set to perform regression-based forecasting for the target variable in the next period. Subsequently, the data from the next period is included in the training set, and the target variable for the following period is forecasted. This iterative process is repeated until forecasts are made for all available samples. To ensure a sufficient sample size for regression fitting ($n > 30$), the rolling pseudo out-of-sample forecasting begins in January 2021. To evaluate the extent to which national ESG in big data enhances macroeconomic forecasting accuracy, this study calculates and compares the out-of-sample forecasting RMSE for models with and without the inclusion of national ESG in big data. National ESG in big data index was constructed using three methods: the monthly count of relevant reports (IPOF_Num), the average of Goldstein scores (IPOF_Goldstein1), and the total of Goldstein scores (IPOF_Goldstein2). The relative improvement in forecasting accuracy associated with different predictive factors was then determined, as shown in Table 3.

Table (1) demonstrates that the integration of the national ESG in big data index leads to significant variations in the predictive performance of different macroeconomic variables. The short-term forecasting ($\tau = 1$) accuracy for import and export trade exhibited the most significant improvement, with increases of 10.13% and 9.02%, respectively. This result suggests that the national ESG in big data plays a significant role and provides substantial value in South Africa's macroeconomic analysis. South Africa maintains a free trade system, engaging in trade with a diverse range of partners, including developed economies such as the European Union and the United States, as well as emerging markets like China and other countries in Asia and the Middle East. As an open economy, South Africa's trade activities exhibit heightened sensitivity to fluctuations in global economic conditions. Accordingly, the integration of national ESG in big data into import and export trade processes demonstrates superior responsiveness and a more pronounced impact. For GDP and M2, accuracy improvements are evident in both short-to-medium-term ($\tau = 1,3,6$) and long-term forecasts ($\tau = 12$). This finding can be interpreted from two perspectives. First, when national ESG in big data exhibits substantial overlap with the historical information of macroeconomic variables (e.g., GDP and M2), the resulting collinearity may diminish the model's predictive accuracy for out-of-sample data. In long-term forecasting, the predictive value of historical information from distant predictors becomes relatively limited. In this context, national ESG in big data assumes a specialized and critical role. On the other hand, national ESG in big data not only directly affects the short-term fluctuations of GDP and M2 but also indirectly influences the long-term trajectory of the macroeconomy and the evolution path of M2 by adjusting market expectations. National ESG in big data plays a particularly significant role in shaping economic expectations, especially given South Africa's history of frequent economic fluctuations. By shaping public opinion, it indirectly influences long-term GDP and M2 through its impact on market confidence, consumer behavior, capital flows, and policy formulation.

The mid-term ($\tau = 3,6$) forecasting accuracy of CPI is markedly higher than its short-term forecasting accuracy ($\tau = 1$). This suggests that national ESG in big data may exert its influence on CPI primarily through indirect mechanisms or with a delayed response. In South Africa, CPI is shaped by multiple factors, including exchange rate volatility and shifts in import and export prices. These factors often become evident over time under the influence of national ESG in big data. The impact of national ESG in big data on enhancing the predictive accuracy of PPI is relatively modest. This suggests that its influence on PPI may be limited or that the signals it provides contain a significant degree of noise. The potential reason may be that the PPI is predominantly influenced by production costs, raw material prices, and the state of the domestic supply chain. In comparison, national ESG in big data appears to impact the PPI indirectly, primarily through exchange rate fluctuations and changes in international commodity prices. The indirect nature of this influence, coupled with the noise effect of public sentiment, results in the national ESG in big data having a less pronounced effect on improving the precision of PPI measurements.

In summary, based on the analysis of the predictive performance of various macroeconomic variables, national ESG in big data demonstrates a significant role in forecasting South Africa's macroeconomic trends. However, its effectiveness in enhancing predictions varies across different macroeconomic indicators. The integration of national ESG in big data has significantly enhanced the accuracy of predictions for South Africa's trade balance, GDP, and CPI. To ensure the robustness of the results, the national ESG in big data index was developed using three distinct approaches and integrated into the LSTM model. The prediction results from columns (2) and (3) were compared with those from column (1), showing general consistency. This consistency further confirmed the robustness of the model's predictive outcomes.

Table 3. Comparison of Predictive Accuracy Improvement of the LSTM Model with the Addition of the National ESG in Big Data Index

target variable	Predicted number of steps τ	IPOF_Num	IPOF_Goldstein1	IPOF_Goldstein2
		(1)	(2)	(3)
GDP	1	-3.43	18.73	-4.22
	3	6.23	-10.16	-1.97
	6	-2.71	-52.71	-15.12
	12	22.03	14.69	9.72
CPI	1	0	18.33	1.67
	3	29.09	27.27	32.73
	6	16.33	13.27	17.35
	12	-3.70	-6.17	7.41
M2	1	0.47	-0.98	3.73
	3	15.21	-3.42	11.41
	6	32.19	8.75	26.88
	12	21.05	27.86	18.58
Imp	1	10.13	7.67	13.61
	3	-10.85	3.22	-1.27
	6	10.22	2.00	1.68
	12	9.68	2.21	18.46
Exp	1	9.02	9.21	9.39
	3	-2.64	2.48	3.52
	6	-5.02	-7.67	-6.30
	12	0.45	-3.36	-8.18
PPI	1	0.38	-1.89	1.89
	3	2.27	-1.52	-1.52
	6	-1.10	-0.55	12.29
	12	0.37	0	2.04

Note: This table is based on the test set from January 2021 to April 2024 and presents the improvement ratio calculated using the RMSE of rolling out-of-sample prediction errors.

Table 4 further validates the contribution of national ESG in big data to enhancing the accuracy of short-term forecasting ($\tau = 1, 2, 3$). Column (1) reports the results of the LSTM model excluding the national ESG, while Columns (2) to (4) present the results of models that integrate the national ESG. In summary, the incorporation of national ESG in big data results in a reduction in cumulative RMSE values, with Columns (2) to (4) showing consistently lower values compared to Column (1). This result provides strong evidence that national ESG in big data plays a pivotal role in short-term macroeconomic forecasting, further substantiating its effectiveness in enhancing the accuracy of macroeconomic trend predictions.

Table 4. Short-Term Forecast Analysis of the LSTM Model with National ESG in Big Data Index

target variable	Predicted number of steps τ	Not included	IPOF_Num	IPOF_Goldstein1	IPOF_Goldstein2
		(1)	(2)	(3)	(4)
GDP	1	3.79	3.92	3.08	3.95
	2	3.97	3.14	3.37	3.69
	3	3.05	2.86	3.36	3.11
Cumulative RMSE		10.81	9.92	9.81	10.75
CPI	1	0.6	0.6	0.49	0.59
	2	0.88	0.73	0.77	0.65
	3	1.1	0.78	0.8	0.74
Cumulative RMSE		2.58	2.11	2.06	1.98
M2	1	2.63	2.23	2.72	2.33
	2	2.7	2.29	2.74	2.4
	3	3.2	2.17	2.92	2.34
Cumulative RMSE		8.53	6.69	8.38	7.07
Imp	1	12.64	11.36	11.67	10.92
	3	11.79	11.27	11.63	11.76
	6	11.8	13.08	11.42	11.95
Cumulative RMSE		36.23	35.71	34.72	34.63
Exp	1	32.69	29.74	29.68	29.62
	2	32.2	32.54	31.46	34.32
	3	31.84	32.68	31.05	30.72
Cumulative RMSE		96.73	94.96	92.19	94.66
PPI	1	5.29	5.27	5.39	5.19
	2	5.12	5.22	5.33	5.2
	3	5.28	5.16	5.36	5.36
Cumulative RMSE		15.69	15.65	16.08	15.75

Note: This table is based on the test set from January 2021 to April 2024 and calculates the rolling out-of-sample forecast error RMSE for different macroeconomic forecasting models. The cumulative RMSE is the sum of the forecast RMSE.

To provide a clearer visualization of the impact of national ESG in big data on macroeconomic forecasting, this paper uses GDP growth rate and CPI growth rate as examples. Figures 4 and 5 present a visual comparison of the fitting sequences generated by the LSTM model without national ESG in big data (LSTM) and the LSTM model incorporating national ESG in big data (LSTM + IPOF_Num) against the actual sequences of the target variables (GDP and CPI).The curves closely align, indicating that both models provide a good fit for the GDP growth rate and CPI growth rate.

With the gradual relaxation of pandemic controls, combined with an economic multiplier effect, South Africa achieved a 19.36% year-on-year GDP growth in the second quarter of 2021, marking a significant turning point in economic growth. However, this abrupt change during the pandemic has introduced new challenges and limitations to macroeconomic forecasting. Prediction models rely on historical data, yet traditional forecasting indicators struggle to capture the effects of sudden events. As shown in Figure 4, the LSTM model shows a considerable lag in recognizing this turning point. While the enhanced LSTM+ESG model, which incorporates the National ESG in Big Data Index, still exhibits some lag, it shows an improvement in overall response time. This suggests that incorporating the National ESG in Big Data Index may help the predictive model more effectively identify and capture the impacts of sudden events, thereby enhancing the model's macroeconomic forecasting capabilities.

Figure 4. Comparison of GDP growth rate forecast fitting effects

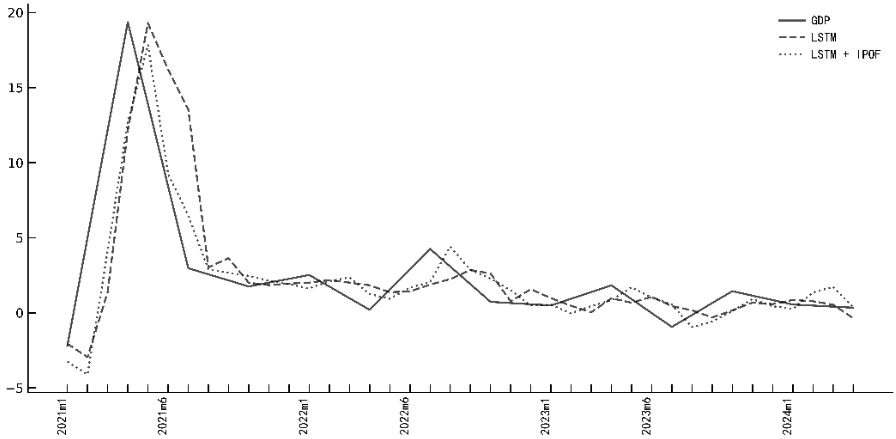
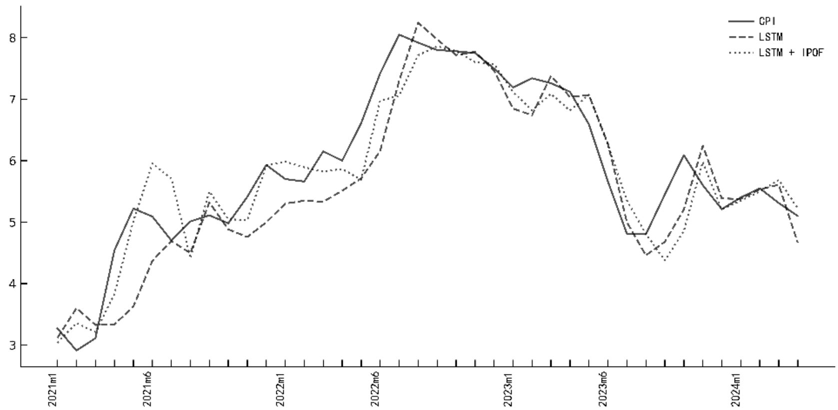


Figure 5. Comparison of CPI growth rate forecast fitting effects



4.4. Expansion Analysis

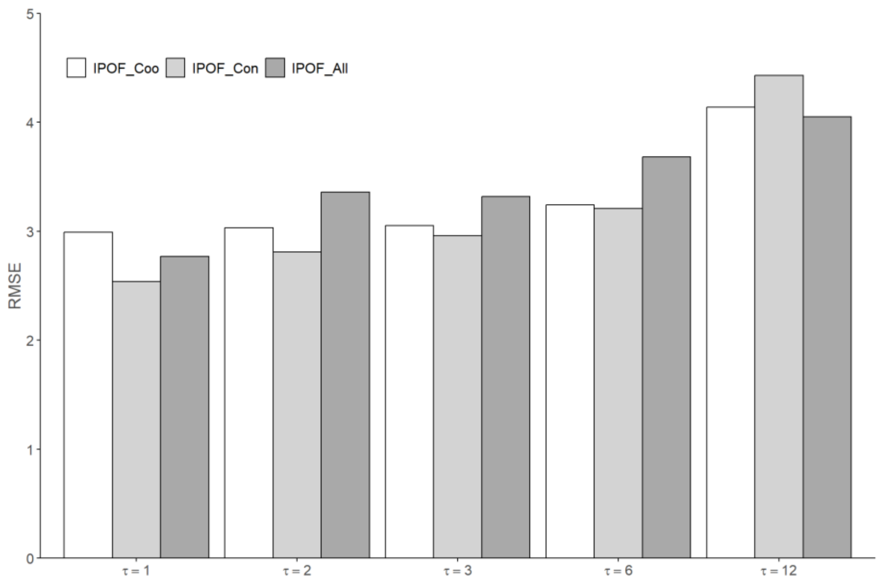
Developing different types of national ESG in big data can enrich the dimensions available for macroeconomic forecasting and enhance model accuracy. At the national governance level, this paper categorizes the national ESG big data index (IPOF_Num) into cooperation (IPOF_Coo), conflict (IPOF_Con), and a combination of conflict and cooperation (IPOF_ALL). Using South Africa's GDP as a case study, it conducts a comparative analysis of the predictive performance across different categories and time horizons.

First, in terms of forecast horizons, South Africa's macroeconomic forecasts perform well in the short and medium term ($\tau = 1, 2, 3, 6$), but the root mean square error (RMSE) increases in the long term. This trend applies not only to GDP forecasts but also to other macroeconomic variables. This phenomenon may be attributed to greater uncertainty and complexity in external conditions associated with long-term forecasting.

Secondly, for different types of National ESG in Big Data, the model's root mean square error is slightly higher when the short-term international cooperation ESG (IPOF_Coo) is included compared to the model with the international conflict ESG (IPOF_Con). This may be because the short-term cooperation ESG index is more affected by temporary external factors, resulting in greater volatility, which makes it harder for the prediction model to accurately fit these fluctuations and leads to a slightly higher RMSE. In long-term forecasting, the RMSE value of national ESG in big data focused on cooperation (IPOF_Coo) is lower than that of national ESG big data associated with conflict (IPOF_Con). On one hand, international conflicts are often sudden and highly volatile, which makes long-term prediction challenging. On the other hand, the cooperation ESG index tends to provide a more stable and lasting influence in the long term. Cooperative relationships usually have longer durations, enabling the model to better identify and fit this stable pattern in long-term forecasts, resulting in a lower root mean square error (RMSE).

Overall, the root mean square error (RMSE) of the conflict and cooperation category (IPOF_ALL) tends to be higher than that of either the conflict (IPOF_Con) or cooperation (IPOF_Coo) categories. Compared to individual conflict or cooperation factors, the combined international sentiment factor is influenced by a broader range of elements, including sudden events, policy changes, economic data, and other global factors, which adds complexity to the prediction model.

Figure 6. Forecast Comparison of Different Types of National ESG in Big Data (GDP)



5. Conclusions

This study aims to explore whether AI-driven National ESG in Big Data can improve the accuracy of South Africa's macroeconomic predictions. Firstly, drawing on the characteristics of South Africa's macroeconomy and existing research findings, this paper integrates a macroeconomic forecasting indicator pool tailored to South Africa's economic features with the national ESG index based on big data to construct econometric models (VAR model, FAVAR model), machine learning models (SVR model), and deep learning models (LSTM model).

Secondly, this study compares the root mean square error (RMSE) of four macroeconomic forecasting models for 1-period, 3-period, 6-period, and 12-period forecasts, identifying the LSTM model as the most suitable for forecasting South Africa's macroeconomy. Subsequently, this paper compares the LSTM models including and excluding the national ESG in big data index and examines the impact of the national ESG in big data on improving forecasting accuracy. Finally, at the national governance level, this paper categorizes the national ESG in big data index (IPOF_Num) into cooperation (IPOF_Coo), conflict (IPOF_Con), and a combination of conflict and cooperation (IPOF_ALL). Taking South Africa's GDP as a case study, it compares the predictive performance across different categories and time horizons. Next, using news data from GDELT event big data, the study develops three versions of the National ESG in Big Data indicator. Finally, an LSTM model incorporating these National ESG indicators is constructed to assess the impact of National ESG Big Data on forecast accuracy.

Based on the empirical results, we draw the following conclusions. (1) When comparing different forecasting models for South Africa's macroeconomy, the deep learning model (LSTM) performs best on the test set, outperforming both traditional econometric and basic machine learning models. (2) The addition of the National ESG Big Data indicators constructed from the GDELT database can significantly improve the accuracy of South Africa's macroeconomic forecasts, though the degree of accuracy improvement varies across different macroeconomic indicators. For South Africa's import and export trade, short-term forecast accuracy improves more noticeably than in the medium and long term. For GDP and M2, accuracy improvements are evident in both short-to-medium-term and long-term forecasts. CPI forecast accuracy shows significant improvement in the medium and long term, while the improvement for PPI is relatively modest. (3) The primary contribution of the National ESG in Big Data to macroeconomic forecasting is its ability to improve short-term forecasting accuracy and enhance the capture of sudden events. This enables quicker reflection of market dynamics and changes in expectations, offering more timely insights for economic policy and investment decisions.

Reference

Ashwin, J., Kalamara, E. and Saiz, L., 2024. Nowcasting euro area GDP with news sentiment: a tale of two crises. *Journal of Applied Econometrics*, 39(5), pp.887-905. https://doi.org/10.1002/jae.3057.

Balcilar, M., Gupta, R. and Kotzé, K., 2015. Forecasting macroeconomic data for an emerging market with a nonlinear DSGE model. *Economic Modelling*, 44, pp.215-228. https://doi.org/10.1016/j.econmod.2014.10.008.

Baker, S.R., Bloom, N. and Davis, S.J., 2016. Measuring economic policy uncertainty. The quarterly journal of economics, 131(4), pp.1593-1636. https://doi.org/10.1093/qje/qjw024.

Barrow, D.K. and Crone, S.F., 2016. A comparison of AdaBoost algorithms for time series forecast combination. *International Journal of Forecasting*, 32(4), pp.1103-1119. https://doi.org/1.10.1016/j.ijforecast.2016.01.006.

Bernanke, B.S., Boivin, J. and Eliasz, P., 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, 120(1), pp.387-422. https://doi.org/10.1007/s00181-013-0702-9.

Bratu, M. and Nicula, A.S., 2024. Sentiment analysis as an innovation in inflation forecasting in Romania. *Marketing i menedžment innovacij*, 15(2), pp.13-25.

Capelle-Blancard, G., Crifo, P., Diaye, M.A., Oueghlissi, R. and Scholtens, B., 2019. Sovereign bond yield spreads and sustainability: An empirical analysis of OECD countries. *Journal of Banking & Finance*, 98, pp.156-169. https://doi.org/10.1016/j.jbankfin.2018.11.011.

Coniglio, N.D., Peragine, V. and Vurchio, D., 2023. The effects of refugees' camps on hosting areas: Social conflicts and economic growth. *World Development*, 168, 106273. https://doi.org/10.1016/j.worlddev.2023.106273.

Consoli, S., Pezzoli, L.T. and Tosetti, E., 2021. Emotions in macroeconomic news and their impact on the european bond market. *Journal of International Money and Finance*, 118, 102472. https://doi.org/10.1016/j.jimonfin.2021.102472.

Elshendy, M., Colladon, A.F., Battistoni, E. and Gloor, P.A., 2018. Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, 44(3), pp.408–421. https://doi.org/10.1177/0165551517698298.

Gentzkow, M., Kelly, B. and Taddy, M., 2019. Text as data. *Journal of Economic Literature*, 57(3), pp.535-574.

Godahewa, R., Webb, G.I., Schmidt, D. and Bergmeir, C., 2023. SETAR-Tree: a novel and accurate tree algorithm for global time series forecasting. *Machine Learning*, 112(7), pp.2555-2591.https://doi.org/10.1007/s10994-023-06316-x..

Gupta, R. and Kabundi, A., 2011. A large factor model for forecasting macroeconomic variables in South Africa. *International Journal of Forecasting*, 27(4), pp.1076-1088. https://doi.org/10.1016/j.ijforecast.2010.10.001.

Gupta, R. and Steinbach, R., 2013. A DSGE-VAR model for forecasting key South African macroeconomic variables. *Economic Modelling*, 33, pp.19-33. https://doi.org/10.1016/j.econmod.2013.03.012.

Gu, S., Kelly, B. and Xiu, D., 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), pp.2223–2273, https://doi.org/10.1093/rfs/hhaa009

Hansen, P.R. and Timmermann, A., 2015. Comment on: Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. *Journal of Business & Economic Statistics*, 33(1), pp.17–21.

Hansson, J., Jansson, P. and Löf, M., 2005. Business survey data: Do they help in forecasting GDP growth?. *International Journal of Forecasting*, 21(2), pp.377-389. https://doi.org/10.1016/j.ijforecast.2004.11.003.

Han, Y., Tian, Y., Yu, L. and Gao, Y., 2023. Economic system forecasting based on temporal fusion transformers: Multi-dimensional evaluation and cross-model comparative analysis. *Neurocomputing*, 552, 126500. https://doi.org/10.1016/j.neucom.2023.126500.

Hochreiter, S. and Schmidhuber, J., 1996. LSTM can solve hard long time lag problems. *Advances in neural information processing systems*, 9.

Hübel, B., 2022. Do markets value ESG risks in sovereign credit curves?. *The quarterly review of economics and finance*, 85, pp.134-148. https://doi.org/10.1016/j.qref.2020.11.003.

Higgins, P., Zha, T. and Zhong, W., 2016. Forecasting China's Economic Growth and Inflation. *China Economic Review*, 41, pp.46–61. https://doi.org/10.1016/j.chieco.2016.07.011.

Hong, Y., Jiang, F., Meng, L. and Xue, B., 2024. Forecasting Inflation Using Economic Narratives. *Journal of Business & Economic Statistics*, (just-accepted), pp.1-27. https://doi.org/10.1080/07350015.2024.2347619.

Jiang, P.C., Feng, G.F. and Yang, H.C., 2022. New measurement of sovereign ESG index. *Innovation and Green Development*, 1(2), 100009. https://doi.org/10.1016/j.igd.2022.100009.

Kelly, B., Manela, A. and Moreira, A., 2021. Text selection. *Journal of Business & Economic Statistics*, 39(4), pp.859-879. https://doi.org/10.1080/07350015.2021.1947843.

Kabundi, A. and Mbelu, A., 2021. Estimating a time-varying financial conditions index for South Africa. *Empirical Economics*, 60(4), pp.1817-1844. https://doi.org/10.1007/s00181-020-01844-0.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G. and Kapadia, S., 2022. Making text count: economic forecasting using newspaper text. *Journal of Applied Econometrics*, 37(5), pp.896-919.https://doi.org/10.1002/jae.2907.

Kohns, D. and Bhattacharjee, A., 2023. Nowcasting growth using google trends data: A bayesian structural time series model. *International Journal of Forecasting*, 39(3), pp.1384-1412.

Liu, T. and Cheng, H., 2024. Can the Classical Economic Model Improve the Performance of Deep Learning? A GDP Forecasting Example. *Romanian Journal of Economic Forecasting*, 27(2), pp.87-110.

Lin, W. and Wei, Y., 2024. Economic forecasting with big data: A literature review. *Journal of Management Science and Engineering,* vol(issue), pp.. https://doi.org/10.1016/j.jmse.2024.01.003.

Lin, J., Fan, J., Zhang, Y., and Chen, L., 2023. Real-time macroeconomic projection using narrative central bank communication. *Journal of Applied Econometrics*, 38(2), pp.202-221.https://doi.org/10.1002/jae.2946.

Larsen, V.H. and Thorsrud, L.A., 2019. The value of news for economic developments. *Journal of Econometrics*, 210(1), pp.203-218. https://doi.org/10.1016/j.jeconom.2018.11.013.

Lupu, R., Calin, A.C., Lupu, I. and Croicu, A. E., 2024. Bubble or not? A time series classification problem using machine learning. *Romanian Journal of Economic Forecasting*, 27(1), pp.5-25.

Naomi, P. and Akbar, I., 2021. Beyond sustainability: Empirical evidence from OECD countries on the connection among natural resources, ESG performances, and economic development. *Economics & Sociology*, 14(4), pp.89-106. https://doi.org/10.14254/2071-789x.2021/14-4/5.

Parikh, A., Kumari, D., Johann, M. and Mladenović, D., 2023. The impact of environmental, social and governance score on shareholder wealth: A new dimension in investment philosophy. *Cleaner and Responsible Consumption*, 8, 100101. https://doi.org/10.1016/j.clrc.2023.100101.

Qin, M., Wan, Y., Dou, J. and Su, C.W., 2024. Artificial Intelligence: Intensifying or Mitigating Unemployment?. *Technology in Society*, vol, 102755.

Shen, S., Xia, L., Shuai, Y., and Gao, D. 2022. Measuring news media sentiment using big data for Chinese stock markets. *Pacific-Basin Finance Journal*, 74, 101810. https://doi.org/10.1016/j.pacfin.2022.101810.

Siami-Namini, S., Tavakoli, N. and Namin, A.S., 2018. *A comparison of ARIMA and LSTM in forecasting time series*. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA), pp.1394-1401. IEEE.

Sims, C.A.,1980. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pp.1-48. https://doi.org/10.1016/j.jmacro.2010.09.008.

Su, C.W., Lv, S., Qin, M. and Norena-Chavez, D., 2024. Uncertainty and Credit: The Chicken or the Egg Causality Dilemma. *Emerging Markets Finance and Trade*, pp.1-19. https://doi.org/10.1080/1540496x.2024.2318299.

Sun, X., Gao, J., Liu, B. and Wang, Z., 2021. Big data-based assessment of political risk along the belt and road. *Sustainability*, 13(7), 3935. https://doi.org/10.3390/su13073935.

Thorsrud, L.A., 2020. Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 38(2), pp.393-409.

Voukelatou, V., Miliou, I., Giannotti, F. and Pappalardo, L., 2022. Understanding peace through the world news. *EPJ data science*, 11(1), pp.2. https://doi.org 10.1140/epjds/s13688-022-00315-z.

Umar, Z., Kenourgios, D. and Papathanasiou, S., 2020. The static and dynamic connectedness of environmental, social, and governance investments: International evidence. *Economic Modelling*, 93, pp.112-124. https://doi.org/10.1016/j.econmod.2020.08.007.

Wei, Y., Wang, Y. and Huang, D., 2010. Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32(6), pp.1477-1484. https://doi.org/10.1016/j.eneco.2010.07.009.

Yemba, B.P., Otunuga, O.M., Tang, B. and Biswas, N.,2023. Nowcasting of the short-run euro-dollar exchange rate with economic fundamentals and time-varying parameters. *Finance Research Letters*, 52, 103571.

Zhang, D., Zhao, Z. and Lau, C. K. M., 2022. Sovereign ESG and corporate investment: New insights from the United Kingdom. *Technological Forecasting and Social Change*, 183, 121899.https://doi.org/10.1016/j.techfore.2022.121899.

Zheng, T., Fan, X., Jin, W. and Fang, K., 2024. Words or numbers? Macroeconomic nowcasting with textual and macroeconomic data. *International Journal of Forecasting*, 40(2), pp.746-761.https://doi.org/10.1016/j.ijforecast.2023.05.006.