Development of Public Dynamic Spatio-Temporal Monitoring and Analysis Tool of Supply Chain Vulnerability, Resilience, and Sustainability

September 2024

A Research Report from the Pacific Southwest Region University Transportation Center

Miguel Jaller, Associate Professor, Department of Civil and Environmental Engineering, Co-Director Sustainable Freight Research Program, Institute of Transportation Studies, University of California, Davis

Juan C. López, Graduate Student Researcher, Department of Civil and Environmental Engineering, University of California, Davis

Alan Jenn, Assistant Professor, Department of Civil and Environmental Engineering, University of California, Davis

Pacific Southwest Region UTC

University Transportation Center

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
PSR-22-46	N/A	N/A

4. Title and Subtitle Development of Public Dynamic Spatio-Temporal Monitoring and Analysis Tool of Supply Chain Vulnerability, Resilience, and Sustainability	**5. Report Date** September 2024 **6. Performing Organization Code** N/A
7. Author(s) Miguel Jaller, Ph.D., 0000-0003-4053-750X Alan Jenn, 0000-0003-4232-0697	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address University of California, Davis Institute of Transportation Studies 1605 Tilia Street, Suite 100 Davis, CA 95616	**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109 [Caltrans TO 070]
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590	**13. Type of Report and Period Covered** Final report (July 2023 – August 2024) **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes
https://doi.org/10.25554/8a5s-dd40

16. Abstract
Supply chains play a pivotal role in driving economic growth and societal well-being, facilitating the efficient movement of goods from producers to consumers. However, the increasing frequency of disruptions caused by geopolitical events, pandemics, natural disasters, and shifts in commerce poses significant challenges to supply chain resilience. This report discusses the development of a dynamic spatio-temporal monitoring and analysis tool to assess supply chain vulnerability, resilience, and sustainability. Leveraging news data, macroeconomic metrics, inbound cargo data (for sectors in California), and operational conditions of California's highways, the tool employs Natural Language Processing (NLP) and empirical regression analyses to identify emerging trends and extract valuable information about disruptions to inform decision-making. Key features of the tool include sentiment analysis of news articles, topic classification, visualization of geographic locations, and tracking of macroeconomic indicators. By integrating diverse and dynamic data sources (e.g., news articles) and using empirical and analytical techniques, the tool offers a comprehensive framework to enhance our understanding of supply chain vulnerabilities and resilience, ultimately contributing to more effective strategies for decision-making in supply chain management. The dynamic nature of this tool enables continuous monitoring and adaptation to evolving conditions, thereby enhancing the analysis of resilience and sustainability in global supply chains.

17. Key Words	18. Distribution Statement
sustainability, supply chain	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	50	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the U.S. Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education, and technology transfer to *improve the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high-growth areas.

Acknowledgments

This study was partially or entirely funded by a grant from the Pacific Southwest Region University Transportation Center (UTC), supported by the California Department of Transportation (Task Order 70) through the University Transportation Centers program. The authors would like to thank the PSR, Caltrans, and the USDOT for their support of university-based research in transportation and especially for the funding provided in support of this project.

Table of Contents

Table of Figures

Development of Dynamic Spatio-Temporal Monitoring and Analysis Tool of Supply Chain Vulnerability, Resilience, and Sustainability

Abstract

Supply chains play a pivotal role in driving economic growth and societal well-being, facilitating the efficient movement of goods from producers to consumers. However, the increasing frequency of disruptions caused by geopolitical events, pandemics, natural disasters, and shifts in commerce poses significant challenges to supply chain resilience. This report discusses the development of a dynamic spatio-temporal monitoring and analysis tool to assess supply chain vulnerability, resilience, and sustainability. Leveraging news data, macroeconomic metrics, inbound cargo data (for sectors in California), and operational conditions of California's highways, the tool employs Natural Language Processing (NLP) and empirical regression analyses to identify emerging trends and extract valuable information about disruptions to inform decision-making. Key features of the tool include sentiment analysis of news articles, topic classification, visualization of geographic locations, and tracking of macroeconomic indicators. By integrating diverse and dynamic data sources (e.g., news articles) and using empirical and analytical techniques, the tool offers a comprehensive framework to enhance our understanding of supply chain vulnerabilities and resilience, ultimately contributing to more effective strategies for decision-making in supply chain management. The dynamic nature of this tool enables continuous monitoring and adaptation to evolving conditions, thereby enhancing the analysis of resilience and sustainability in global supply chains.

Development of Dynamic Spatio-Temporal Monitoring and Analysis Tool of Supply Chain Vulnerability, Resilience, and Sustainability

Introduction

Supply chains play a critical role in driving global and national economic growth. They facilitate the efficient movement of goods from producers to consumers, thereby benefiting society as a whole. However, geopolitical events, pandemics, natural disasters, and changes in commerce over recent decades have increasingly disrupted, temporarily or permanently, supply chains. These disruptions have significantly impacted the economy and society, reducing sales, supply chain bottlenecks, labor and material shortages, and decreased production capacities. As a result, consumers face higher prices and longer delivery times for essential goods. Research studies have estimated that over 56% of companies suffer a supply chain disruption annually globally [1].

These disruptions can stem from various factors, from anthropogenic to natural and geopolitical causes. Statistics reveal that between 1980 and 2022, natural disasters in the U.S. alone resulted in approximately $1.2 trillion in damages, with their frequency rising each decade. Furthermore, these events have led to the closure of about 40 to 60% of small businesses affected by such disasters [1]. In response to these challenges, the U.S. government has formulated policies through Presidential Executive Orders and initiatives by the Department of Transportation (DOT) to identify and address vulnerabilities and bolster supply chain resilience, recognizing their crucial importance to the overall economy [2–6].

However, a significant obstacle in addressing supply chain disruptions lies in the lack of publicly accessible data, exacerbated by the dynamic nature of supply chains. Previous research has underscored the effectiveness of using news data to identify and quantify the impacts of disruptions on supply chains over time [7]. Therefore, news articles and other sources can be valuable for tracking and evaluating the various events influencing supply chains, offering quasi-real-time insights into supply chain developments impacting macroeconomic indicators such as inflation, employment, and resource shortages.

This translational project developed a dynamic spatio-temporal monitoring and analysis tool to assess supply chain vulnerability, resilience, and sustainability. This tool collects, processes, and analyzes descriptive and quantitative data to identify and project relevant disruptions in the different segments of the supply chains, as well as anticipate potential impacts and changes on transportation logistics. It employs text analysis and regression modeling to extract information from news and other sources, translating it into actionable data to develop empirical econometric models. These models help discern the magnitude and impact of different factors on the resilience and sustainability of supply chains. This data will generate public and dynamic information showcased in a dashboard that supports the development of planning and management strategies that can anticipate future disruptions and promote the economic and environmental improvement of the sector.

The project leveraged previous research conducted by the team and involved the following research tasks:

Task 1: Data collection and analysis. This task involved identifying additional data sources to complement the team's previous work, including the U.S. Bureau of Labor Statistics, trade information, and other census products. As part of this task, the team developed scripts to connect and process the information from those sources and enhance the data collection process through different APIs. The proposed data collection methodology gathers information automatically, conducts data cleaning, and generates an actionable database.

Task 2: Update data processing, text analysis, and relationship modeling tools. The team worked on automating the processes to perform content/topic, relationship framework, and spatio-temporal analyses in the tool. In this task, the team performed a validation test for the tool. Additionally, the team generated an initial set of performance metrics to analyze.

Task 3: Dashboard creation, testing, demonstration, and incorporation of feedback. This task involved the development of a dashboard to display and visualize the different metrics and results.

This report discusses the work conducted on the various tasks. Section 2 provides background on supply chain disruptions and summarizes key resilience topics in national policies and plans. Sections 3 to 5 discuss the work conducted in Tasks 1 to 3.

Background on Supply Chain Disruptions and Resilience

Supply chains are complex networks encompassing facilities, transportation, and the marketing of goods, involving the production of raw materials, their conversion into intermediate and final products, and their transportation and distribution to customers. These networks comprise various entities working together to ensure the continuous flow of materials, products, services, finances, and information to meet customer demands efficiently and cost-effectively [8]. Supply chains play a crucial role in driving global and national economies by fostering economic growth by creating well-paying jobs, encouraging innovation, and contributing to overall economic size and performance [9]. In 2015 alone, supply chains employed over 44 million people in the U.S., representing about 37% of all industries' employment with wages higher than the average. They also heavily focus on Science, Technology, Engineering, and Math (STEM) jobs, indicating innovation potential [10].

Despite their significant economic contribution, supply chains are among the most vulnerable to disruptive events, including economic downturns, geopolitical tensions, and natural disasters. Research indicates that supply chains were susceptible to economic downturns such as the dotcom burst (2000-2002) and the Great Recession (2007-2009), resulting in a decline in annual employment growth (about -10%) compared to national averages during those periods (-3%) [10]. Additionally, disruptions such as shortages of raw materials, containers, or truckers, transportation and logistics issues, pandemics, and natural disasters can significantly impact supply chains by causing product scarcity, sales losses, revenue declines, and increased material and transportation costs, leading to inflationary pressure, delays, and decreased productivity [1,11].

Recent research has highlighted the role of supply chain disruptions in inflation generation across countries, including the Euro area and the U.S. [12]. During the COVID-19 pandemic, supply chain disruptions were an important source of inflation in the U.S. [13]. The transportation and logistics sector experienced disruptions, resulting in delays at ports and, coupled with elevated consumer demand, led to a surge in container spot shipping prices to more than 1000% of 2019 levels, contributing to rises in both Producer and Consumer Price Indexes [3]. McKinsey estimates that the cumulative damages of the disruptions an industry can experience over a decade can equal almost 45% of one year's profit [14]. These impacts at the firm level propagate into macro-level effects, driving prices up and inducing inflation [3].

Disruptions can vary in magnitude, generating losses ranging from millions to trillions of dollars [14], and can be classified according to [1]:

a. The related supply chain echelons include production, supply, and transportation.

b. The reasons for the disruption, including natural events, pandemics, hurricanes, terrorism, wars, delays, systems breakdowns (e.g., information breakdown and cyberattacks), inaccurate demand forecasts, intellectual property issues, procurement challenges, receivables, inventory management uncertainties, and capacity constraints [14,15].

c. Their frequency of occurrence, with common cyberattacks being more frequent than pandemics.

d. Their nature and source spanning process, control, demand, supply, and environmental risks.

e. The entities they affect.

Supply chain resilience is critical for safeguarding the economy, ensuring business continuity, reducing financial losses, and maintaining competitiveness. A resilient supply chain can withstand, respond to, and recover from disruptions [16–19]. Unlike the reactive approach of traditional supply chain management to disruptions, supply chain resilience adopts a proactive stance to increase business continuity during crises and minimize adverse impacts on the economy [1]. Achieving resilience entails better data management, investments in redundancy, improved communication among stakeholders, and an increased ability to substitute inputs and transportation channels [1,3].

Various strategies contribute to supply chain resilience [20], with the most cited ones including increasing flexibility, creating redundancy, fostering collaborative relationships, and enhancing supply chain agility. Current policies aim to support these strategies by investing in trade corridors, gateways, and critical infrastructure and strengthening governance to address prevailing supply chain vulnerabilities.

Table 1 summarizes the most relevant policies in the U.S., including references to supply chain resilience.

Most of these policy efforts focus on identifying and addressing existing vulnerabilities while enhancing supply chain resilience by enabling faster response and recovery from disruptions. However, the absence of high-frequency predictive models and publicly accessible data limits the ability to continuously track and monitor global, national, and local supply chain events. This constraint hinders the ability to anticipate potential impacts and make proactive decisions to mitigate the impact of such disruptive events. Therefore, this project developed a public, dynamic spatio-temporal monitoring and analysis tool to evaluate supply chain vulnerability, resilience, and sustainability. Moreover, the tool will help support more effective decision-making in supply chain management.

Table 1. Examples of U.S. plans fostering supply chain resilience.

Policy document	Organization	Actions	Ref.
Strategic Plan 2022 - 2026	U.S. DOT	• Addressing critical supply chain vulnerabilities, including issues, risks, and bottlenecks. • Convening stakeholders for commitments, aiming to enhance coordination and cooperation among key actors. • Assessing trends and technologies that can enhance supply chain resilience, including investments, business models, and labor and workforce considerations. • Investing in trade corridors and gateways to support the movement of goods. • Promoting intermodal connections to facilitate the movement of goods among different modes of transportation. • Enhancing freight safety. • Addressing last-mile delivery challenges. • Strengthening domestic sourcing and workforce.	[6]
Supply Chain Assessment of the Transportation Industrial Base: Freight and Logistics	U.S. DOT	• Infrastructure investment in freight projects, domestic manufacturing, intermodal capacity, and intelligent transportation systems. • Provide planning and technical assistance in implementing freight policy across states and MPOs. • Improve research and data. • Provide rules and regulations to support domestic production of critical equipment, promote competition in ocean shipping and rail industries, and grant emergency regulatory relief for supply chain disruptions. • Encourage coordination and partnerships.	[5]
Executive Order 14017 on Securing America's Supply Chains	President of the U.S.	• Conduct reviews on supply chain risks and vulnerabilities in critical sectors. • Annual sectoral assessments are required in defense, energy, transportation, and agriculture. • Recommend actions to improve resiliency, including reshoring supply chains, developing domestic production, cooperating with allies, building redundancy, and potential regulatory/policy changes. • Consider the establishment of a quadrennial supply chain review and monitoring process. • Evaluate the potential diplomatic, economic, trade, and other policy tools to strengthen supply chain cooperation with allies and partners. • Examine the reforms needed in domestic/international trade rules, federal procurement, workforce education, preventing monopolization, and supporting small businesses and underserved communities. • Assess the federal incentives or programs that could attract investment in critical goods and materials production in the U.S.	[4]

Policy document	Organization	Actions	Ref.
Other policy indications	The White House	• Create the new White House Council on Supply Chain Resilience to advance a long-term, government-wide strategy and coordinate agency efforts. • Use the Defense Production Act to boost domestic manufacturing of essential medicines, medical supplies, and critical inputs. • Foster cross-agency supply chain data sharing and monitoring capabilities. • Invest in critical supply chains like clean energy technology, food/agriculture, and defense industrial base. • Develop planning efforts like the quadrennial supply chain review, smart manufacturing plan, and mapping/tracking capacities for supply chain risks. • Deploying centers and exercises focused on supply chain resilience. • Engage with the private sector and public stakeholders. • Strengthen international cooperation. • Promote a greater diversification of suppliers and sources, both domestically and internationally. • Improve supply chain data monitoring and analytical capabilities.	[3,23]
Joint Statement on Cooperation on Global Supply Chains	U.S. Department of State	• Promote transparency through information sharing. • Increase global capacities for multiple, reliable sources of materials, inputs, and goods across priority sectors and logistics infrastructure. • Work with allied countries to promote predictability, openness, fairness, and non-discrimination in economic relations. • Deepen consultations to identify and address risks from supply dependencies and vulnerabilities in critical infrastructure. • Encourage global sustainability, responsible business conduct, and upholding labor rights across supply chains. • Engage businesses, workers, civil society, different levels of government, and other stakeholders.	[22]
Research, development, and technology strategic plan 2022-2026	U.S. DOT	• Develop data and tools to assess freight system performance and support performance-based freight planning and policies. • Identify and promote tools and practices to improve freight system safety, reliability, and resilience.	[21]

Background on Statistical Modeling with Text-based Indicators

The analysis of textual data has become increasingly important in various scientific fields. As a result, using forecasting models based on textual data has gained significant attention in recent years due to their potential to offer real-time insights into macroeconomic indicators and supply chain disruptions. Statistical techniques have played a crucial role in developing robust forecasting models that use textual data from sources like news articles and social media posts.

Various methodologies and statistical techniques are used in the literature to track and predict economic and supply chain activities. Ingle et al. [24] comprehensively review demand forecasting methodologies, including traditional statistical models, machine learning, deep learning, and hybrid models. Barbaglia [25] combines news information with a state-of-the-art machine learning model, specifically an auto-regressive probabilistic Recurrent Neural Network model (DeepAR), to improve the prediction accuracy for economic and financial time series. Joshi [26] used latent semantic analysis to explore the growth of literature in supply chain analysis, particularly in service supply chain systems, from operational to demand-centric forecasting. Ostaysi and Bolturk [27] presented various fuzzy methods for demand forecasting in supply chain management. Their methodology involves fuzzy time series, fuzzy regression, adaptive network-based fuzzy inference systems, and fuzzy rule-based systems for demand forecasting.

Barbaglia [28] evaluated the informational content of news-based sentiment indicators for forecasting gross domestic product and other macroeconomic variables of the five major European economies. The authors collected newspaper articles and computed sentiment indicators to capture the public's overall attitude toward varying aspects of the economy. These news-based sentiment indicators were then integrated into econometric models for nowcasting and forecasting GDP. Feuerriegel and Gordon [29] also presented a methodology for predicting macroeconomic indicators using financial news. To this end, they utilized machine learning models to handle non-linear relationships and high-dimensional data derived from text features. Moreover, the authors applied Lasso regularization methods to prevent overfitting and enhance the model's predictive performance and interpretability. This study established a benchmark based on traditional time series models, such as autoregressive moving average models and their variations, to compare and evaluate the performance of the developed models.

On the other hand, Aprigliano et al. [30] analyzed how text-based indicators can be used to improve the nowcasting and forecasting of the Italian GDP. The authors employed BMA to address model uncertainty when selecting predictive variables. BMA allows for combining information from multiple models to improve prediction accuracy. This approach provided a more accurate and nuanced prediction by accounting for model uncertainty, which is particularly advantageous in situations like economic recessions where high volatility and uncertainty are present. A is a powerful tool for developing high-frequency forecasting models as it addresses the challenge of variable selection in regression models, particularly in settings with many potential regressors and limited observations [31]. Other authors have used this approach to assess supply chain risks [32] and forecast economic activity and macroeconomic indicators such as GDP and inflation [33–35].

Task 1 – Data Collection and Analysis

News and metrics serve as crucial sources of information for understanding and forecasting supply chain dynamics. Leveraging these sources allows for the identification of emerging trends and potential disruptions. However, it is essential to recognize the risks associated with misinformation and fake news, which can introduce uncertainties and impact the reliability of analytical models. Therefore, access to credible news sources is paramount for navigating these challenges effectively.

News Sources

Leveraging news information holds the potential for forecasting significant future disruptions based on emerging trends. However, recent studies have underscored the risks posed by misinformation and fake news [36], which can exacerbate supply chain uncertainties and undermine the reliability of empirical models. Hence, access to credible news sources is essential for effectively navigating these challenges.

In the project, we use two main sources of news articles:

1. **LexisNexis**: LexisNexis, particularly its academic platform NexisUni, stands out as a pivotal information source. NexisUni offers extensive access to historical news archives, comprehensive local and global news coverage, company profiles, executive data, and legal documents [37]. This database has been used in some research studies [7,38]. In this project, initial efforts involved collecting news articles from NexisUni in Word format, which proved laborious and time-intensive. To streamline this activity, the team developed an algorithm using Python 3.12.2 [39] for the automated collection of news articles. This algorithm enables targeted searches based on specific keywords and time ranges, automating the download of Word files containing article contents.
2. **NewsAPI**: NewsAPI is an HTTP REST Application Programming Interface (API) for retrieving live articles from various web sources [40]. It offers headlines and URLs from news sources and blogs across the internet. The team uses the '*newsanchor'* package [41] in R 4.3.3 [42] to use this resource. However, as this API only provides headlines, the team developed a scraping tool using the '*httr'* package [43] to retrieve complete articles.

The news articles are collected based on the following query: (USA or United States) and (supply chain or supply-chain) and (disruption or resilience) and (retailer or warehouse or transportation or factory). Results were limited to news, newspaper, and web news categories, retrieving only articles in English. After retrieval, the news articles are processed, and duplicates and those with incomplete information, such as publication date, are discarded. Table 2 summarizes the number of retrieved and processed news from each source.

Table 2. Collected and processed news by source.

Source	Time Range	News retrieved
NexisUni	January 2018 – Present date[1]	119,035
NewsAPI	August 2023 – Present date[1]	1,247

[1] Collecting news daily – Numbers up to 6/10/2024

Macroeconomic and Other Supply-Chain Related Data

The tool also uses various sources of information to gather insights on the macroeconomic events both at the national and state level, inbound cargo data in California, and operational conditions of state highway systems, as described below:

1. **FRED:** The Federal Reserve Economic Data (FRED) is a public online database that provides state, national, and international economic indicators [44]. This tool, maintained by the Economic Research Division of the Federal Reserve Bank of St. Louis, provides an API that allows users to automate data collection. This database gathers information from various of the most important macroeconomic data sources in the U.S., including the Census Bureau, Bureau of Labor Statistics, and Bureau of Economic Analysis. In this project, the tool performs automated data retrieval from FRED using the *'fredr'* [45] package in R. Our database contains approximately 1270 different datasets with frequencies that go from updates each year, semester, quarter, month, weekly, and daily.

2. **Federal Reserve Bank of New York:** The Federal Reserve Bank of New York [46] publishes the Global Supply Chain Pressure Index (GSCPI), a monthly indicator that assesses the impact of supply constraints on economic outcomes, particularly regarding goods and producer price inflation in the U.S. and the euro area. The project team developed an automated tool to collect the GSCPI data weekly.

3. **Census:** The U.S. Census Bureau, through its international trade division [47], publishes monthly U.S. export and import statistics, which are disaggregated at the port level. This dataset provides information on each U.S. port's monthly values and quantity of imports and exports. This data is automatically gathered via the Census API [48].

4. **Port Optimizer:** An automated data retrieval tool is developed to gather daily metrics from PortOptimizer [49], including average berth time at the Port of Los Angeles (L.A.) and truck queue, terminal, and total turn time at the ports of L.A. and Long Beach (L.B.). These metrics are used to estimate congestion and efficiency of operations at California's two most important ports.

5. **X**: Automated gathering data from the X (former Twitter accounts) of California's Highway Patrol (CHP) [50] and CalFire for tracking disruptions at the highway level in California.

Task 2 – Data Processing, Text Analysis, and Relationship Modeling Tools

The research team implemented a modular processing procedure in this task, illustrated in Figure 1. With this pipeline, we aim to enhance data processing, text analysis, and relationship modeling efficiency and effectiveness. This modular approach allows flexibility and scalability, enabling adaptation to new requirements and data sources. This procedure consists of several components:

1. **Static module:** responsible for setting the lexicons required for sentiment and topic analysis. It remains static throughout the processing.
2. **Dynamic components:** include news collecting for dynamically gathering news articles from the sources, text analytics for extracting the main features of the news, metrics collecting and processing for dynamically obtaining, filtering, and analyzing the macroeconomic indicators, and statistical module for developing the econometric modeling based on the processed news and indicators.
3. **Visualization component:** Corresponds to the dashboard, which provides a graphical, interactive, and dynamic representation of the analyzed news, metrics, and models.

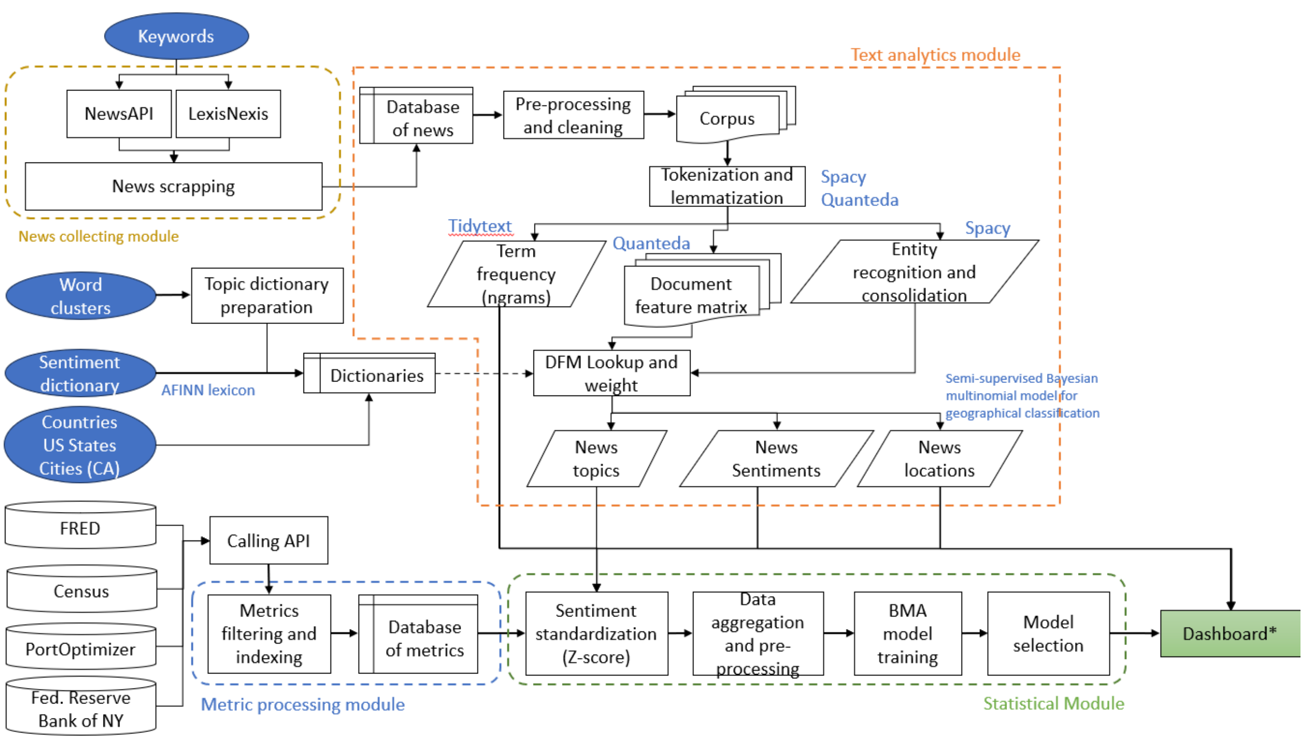


Figure 1. Data processing flow chart.

Static Modules

The static module aims to set the dictionaries for classifying news into topics, sentiments, and locations, thereby setting the foundation for the statistical and visualization modules.

1. **Topics lexicon:** Topics are categorized into eight groups relevant to supply chain risks: political, environmental, financial, supply and demand, logistics, system, and operations. Each topic group was also divided into subtopics based on seed words sourced from [51]. Furthermore, we used the WordNet dictionary through the Wordnet [41] package in R to obtain synonyms for these topics, as we did in [7]. Appendix A – Topics Lexicon presents the lexicon adopted for each topic. The lexicon was adapted from previous research projects [7], a literature review, and key supply chain-related policies outlined in Table 1.

2. **Sentiment lexicon:** Sentiment and emotion analysis of text encompasses a variety of methodologies, including lexicon-based, machine learning-based, and deep learning-based techniques [52]. These methodologies aim to extract semantic information from text using natural language processing (NLP) techniques to discern the writer's attitude, which can be positive, negative, or neutral [53]. Recent research has underscored the effectiveness of lexicon-based techniques in sentiment and emotion analysis, prompting the adoption of this methodology.

   Two primary methods are prominent in the lexicon-based approach: dictionary-based and corpus-based. The dictionary-based method, exemplified by lexicons like ANEW, AFINN, VADER, and SentiwordNet, assigns values to emotion-bearing words, facilitating sentiment analysis [54,55] and allowing for quantitative analyses. We utilize the AFINN [56] lexicon for this project due to its proven performance and accessibility [54]. AFINN comprises over 3300 English words rated for valence on a scale from -5 (negative) to +5 (positive), providing a quantitative framework for sentiment analysis.

3. **Locations lexicon:** Besides topics and sentiments, news articles are also categorized based on geographical locations mentioned within their contents. The classification of locations is conducted at three hierarchical levels: countries, states within the U.S., and cities/municipalities within California. Three comprehensive lists of countries, states, and cities are used to facilitate this classification.

Dynamic Modules

These modules constitute the tool's core functionality. They are responsible for daily queries and retrieval of information, news processing, and metrics retrieval and analysis to ensure the dashboard remains current. They encompass news collecting, text analytics, metrics collecting and processing, and statistical modeling based on the processed data.

News Gathering

This module involves the daily execution of four interconnected algorithms to gather news articles from various sources:

1. **NexisUni downloading:** This algorithm, developed in Python as described in Section 3.1, retrieves search news articles from NexisUni. It is scheduled to execute automatically daily to obtain news published the day before and download them in Word files.
2. **NexisUni processing:** Once news articles are retrieved in Word files, they are converted into a structured format using the LexisNexisTools package in R. This algorithm processes the news articles, filtering out incomplete information such as publishing dates.
3. **NewsAPI:** Developed in R, this algorithm consists of two main steps. First, the API is triggered to perform the query, obtain news headlines and URLs, and filter out duplicates. Second, scrap the articles' content from their respective websites. Due to the multiple sources from which news is retrieved via NewsAPI, the content-collecting algorithm was designed to be general and flexible.

All four algorithms are automated to execute daily and collect the most recent news with a lag of 24 hours. News collected from Algorithms 2 to 4 are stored in database files for improved efficiency and performance in storage. The team uses the RSQLite [57] package in R for this.

Text Analytics

The text analytics module processes the stored news articles to identify n-grams, topics, sentiments, and geographical locations through text mining tasks. The process involves several steps:

1. **Corpus setup:** The collection of news articles is set up as the corpus, requiring various preprocessing procedures. These procedures include eliminating duplicates among news sources, assigning an I.D. to each news article, and removing non-valuable strings such as websites, non-ASCII characters, and excess whitespaces. The corpus is parsed using SpaCy [58], an open-source library for Natural Language Processing (NLP) in Python, which is known for its effectiveness and efficiency [59]. SpaCy is utilized in R through the *'spacyr'* [50] package.
2. **Term frequency:** Initially, the corpus is tokenized and lemmatized to prepare it for analysis. A second cleaning process removes email addresses, numeric tokens, and stopwords. Then, the team estimates each term's frequency using the *'tidytext'* package [51] to tokenize the corpus into a per-term-per-document (unigram) data frame. However, since term frequency alone may not adequately represent word relevance due to Zipf's Law [60], Term Frequency-Inverse Document Frequency (TF-IDF) is adopted. TF-IDF is a popular metric for measuring the relationship and relevance of words in documents, where tokens with higher TF-IDF are

considered more representative [61]. TF-IDF values are computed using the *'tidytext'* package in R for both unigrams and bigrams, based on Eq. 1

$$tf.idf_t = tf_t \cdot \log\left(\frac{N}{n_t}\right) \qquad \textbf{Eq. 1}$$

Where $tf_{t,d}$ is the frequency of the term t, N is the total number of documents in the corpus, and n_t is the number of documents that contain token t.

3. **Entity recognition and consolidation:** Named Entity Recognition (NER) is employed to identify geographical locations mentioned in news articles, focusing specifically on geopolitical entities (GPEs). The *'spacyr'* package supports this task. As NER is usually performed at the token level, i.e., unigrams, multi-word entities are consolidated into single entities to enhance accuracy. This task aims to identify GPEs mentioned in the news, not evaluate their relevance.
4. **Document-Feature Matrix (DFM):** The corpus is converted into a Document-Feature Matrix (DFM) using the *'quanteda'* [62] package in R to evaluate the relevance of GPEs and perform sentiment and topic analysis based on the dictionaries set in the static module. This toolkit operates based on a DFM, which counts the occurrences of tokens in documents for quantitative analysis and allows the management and process of corpus for quantitative analysis to be quick and efficient [55].
5. **Topics and sentiment analysis:** The *'quanteda'* package facilitates the application of dictionaries to the DFM, enabling the identification of features or tokens that align with the values within the dictionary. This process is structured into three distinct stages:
   a. Sentiment analysis – share of words: quantify the occurrence of positive and negative words within each document. Given the varying lengths of the news articles under analysis, we measured the proportion of words with positive or negative connotations based on the sentiment dictionary.
   b. Sentiment analysis – sentiment score: a sentiment score is assigned to each news article based on the frequency of positive or negative words, weighted according to their sentiment scale. Sentiment scores for all articles were standardized from -1 to +1 to ensure better comparability across documents within the corpus.
   c. Topic share: utilizing the topics' lexicon, we calculated the relative frequency with which tokens associated with each topic appeared within the content of a news article. This allows for assessing the prominence of specific topics across the corpus.
   d. Geographical location: performed through a Semi-supervised Bayesian multinomial model [64]. This approach utilizes a seed dictionary (locations dictionary) for semi-supervision and calculates association scores of words based on co-occurrences in the

corpus. It predicts locations associated with documents by finding the largest total scores weighted by the normalized frequency of words in the corpus. This approach has been more precise than simple lexicon matching [65]. A California index was developed to quantify the relevance of the state within the news articles published in a day, *d*, based on Eq. 2:

$$CA_d = \frac{\sum_i^{N_{t,d}} \mathbf{1}[CA\ Article]_i}{N_{t,d}} \qquad \textbf{Eq. 2}$$

Where *i* represents an article published on a day *d*, $N_{t,d}$ indicates the total number of articles published on a day *d*, and the indicator function $\mathbf{1}[CA\ Article]_i$ is a vector that counts when an article mentions California in its contents. This indicator function **1**[x] takes the value of 1 if x is an EPU article and 0 otherwise.

The dynamic module generates four main outcomes from the processed news articles: TF-IDF for unigrams and bigrams, sentiment words and scores of news, topic share, and geographical locations of news. These outcomes are stored as databases for efficiency. In addition, this module is automated to execute daily, processing the news collected each day. The module retrieves the latest news articles daily, processes them to extract relevant information, and generates the desired outcomes. These outcomes are then updated by incorporating the new information.

Metrics Estimation

Automated retrieval of metrics data is conducted weekly, ensuring that the dashboard remains updated with the latest macroeconomic indicators. The process involves different steps based on the data sources:

1. **FRED**: Upon retrieval, datasets from the Federal Reserve Economic Data (FRED) are filtered based on several criteria:

    a. Geographic Region: Focus is placed on datasets relevant to the United States and California.

    b. Frequency: Priority is given to datasets with monthly, weekly, and daily frequencies, ensuring granularity and timeliness of data.

    c. Last Update: Datasets updated within the last quarter or more recently are prioritized, reflecting the still-in-use economic indicators.

    d. Popularity: Datasets with a relevance index over 50 are considered for inclusion in the dashboard.

    The selected FRED datasets are indexed and classified into topics established in the static module. This classification is achieved through a matching process with the lexicons defined earlier. By associating datasets with relevant topics, users can easily access and interpret economic indicators within the context of the managed topics on supply chain risks.

FRED datasets and their corresponding metadata and classifications are stored in .csv files. This format ensures compatibility and accessibility for further analysis and visualization within the dashboard.

2. **Federal Reserve Bank of New York:** Data related to the GSCPI is retrieved automatically and stored in a .csv file to ensure compatibility and accessibility for further analysis and visualization within the dashboard.
3. **Census**: Upon automated retrieval, data is classified into imports and exports and stored in separate .csv files. These files contain the values and quantities of imports and exports traded from each U.S. port, allowing for further analysis at the national or state level.
4. **Port Optimizer**: Once collected, the data is stored in a .csv file containing information for the ports of L.A. and L.B.

Combining these diverse data sources enriched the dashboard with diverse and comprehensive metrics, providing a holistic view of economic and supply chain trends.

Statistical Module

Various statistical models analyze news data to extract relevant information about macroeconomic events. These models primarily fall into two categories: time series [28,29] models employing Lasso penalization techniques [25,66] and Bayesian models [30,33] utilizing methods such as Bayesian Model Averaging (BMA). BMA combines forecasts from multiple models to enhance accuracy and reliability by weighing each model's forecasts based on their relative skill levels [67,68]. Unlike frequentist time series models, BMA offers better interpretability for regressors and predictive densities that incorporate uncertainties related to model selection and coefficient estimation [30].

To leverage news data for forecasting the effects of supply chain disruptions on macroeconomic indicators like inflation, employment, and international trade, the team opted for the BMA technique. BMA allows for a more comprehensive assessment, considering the uncertainty inherent in supply chains and their interactions with the broader economy. Our approach involved several steps:

1. **Sentiment standardization:** Standardizing sentiment scores and words using Z-scores (Mean = 0 and variance = 1) ensures comparability across news articles.
2. **Aggregation and preprocessing**: Text analytics outcomes are aggregated daily and then to a monthly frequency to align with the release frequency of macroeconomic indicators. Additionally, we opted to process the topic share using the results at the subtopic level to improve granularity and prediction performance.

3. **BMA modeling and training:** BMA models were built using the 'BMS' package in R Studio [69], involving the following substeps:

   a. **Preprocessing data.** The data was divided into training and test sets. Training data covers observations from 01/01/2018 to 06/31/2023, while the test set includes observations from 01/07/2023 onwards. Training data facilitated model building and preliminary analyses while testing data was used for performance evaluation. Additionally, some macroeconomic variables to predict, like imports and exports, were seasonally adjusted to have a clearer view of nonseasonal trends.

   b. **Lag testing.** In the data-gathering process, we collected news information in quasi-real time. However, many macroeconomic metrics and indicators are not published with this frequency. These indicators often rely on historical data to assess the economy's behavior in a month. Moreover, the temporal relationship between when the media reports a supply chain disruption and when the economy reacts to it is uncertain.

      Consequently, we conducted a statistical assessment to determine the best lag for predicting each metric. We built and tested three predictive models using 0-, 1-, and 2-month lags, respectively, using training data and evaluating them with testing data. The evaluation was based on the log-predictive score, a widely used metric for gauging a model's predictive performance across its entire distribution [31]. This approach helped us identify the lag that yielded the best predictive performance.

   c. **Variable selection.** Handling topic sharing at the subtopic level increased the complexity of the predictive models, necessitating a larger number of regressors. Consequently, the BMA approach generated a substantial number of models. To streamline this process, we first developed an initial model incorporating the lag identified in the previous step and all variables at the subtopic level. This model served as a baseline for filtering out less relevant variables. We then assessed various thresholds of Posterior Inclusion Probability (PIP) acceptance using the log-predictive score. Generally, we observed that removing variables with a PIP lower than 0.2 significantly enhanced the model's predictive performance, striking a balance between complexity and accuracy. PIP represents the probability of including a specific predictor (variable) in the model [70].

   d. **Model selection and parameters.** Once the lag and variables for inclusion were identified, we proceeded to select a final model using the complete dataset. Subsequently, we generated prediction intervals and assessed the marginal effects of the variables to gauge their impact on the predicted outcome. Furthermore, we conducted tests on the autocorrelation and partial autocorrelation of residuals to detect any spatial autocorrelation.

Prediction models were saved in .rsd objects for efficient communication with the dashboard, while prediction intervals were stored in .csv files for simpler utilization by the dashboard.

Statistical Modeling Implementation

This section presents and discusses the results of the BMA developed to forecast the GSCPI. Results for other macroeconomic indicators are provided in Appendix B – BMA Models. Lag testing indicated that the 1-month lag model best predicts the GSCPI. Also, it was found that removing variables with a PIP lower than 0.2 significantly improves the model's predictive performance, reaching a tradeoff between complexity and accuracy. Figure 2 illustrates the model's PIP, revealing that the final model contains an average of 8.67 variables and comprises more than 2000 models. In the model inclusion graph, red indicates a negative correlation between the regressor and the outcome, blue indicates a positive correlation, and white spaces indicate the variable is not included in that model. The more frequently a variable appears, the more relevant it is for predicting the outcome.

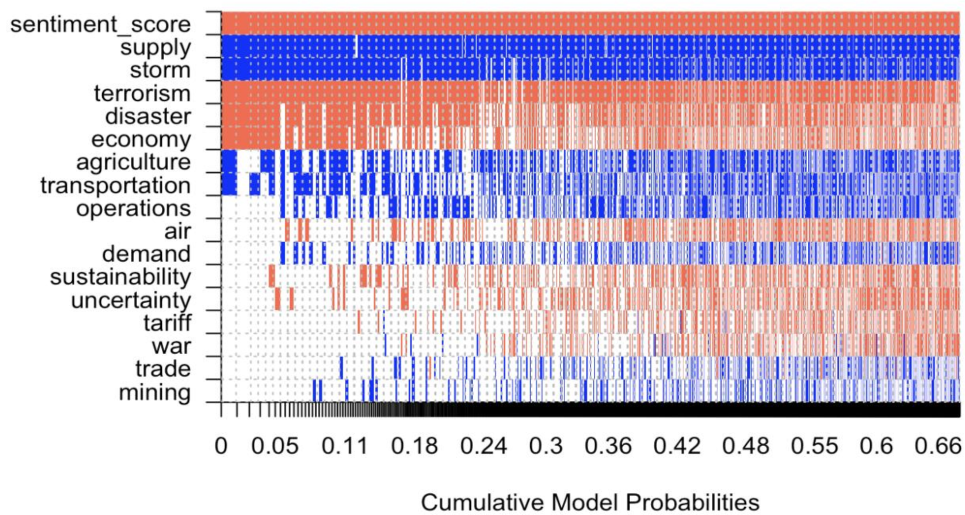


Figure 2. Model inclusion considering the best 2000 models.

From Figure 2, the BMA identifies the sentiment score as the most relevant variable for predicting the GSCPI, with a negative correlation. This suggests that more negative news sentiment corresponds to an increase in the GSCPI, aligning with the GSCPI's role in summarizing potential supply chain disruptions. Additionally, at the subtopic level, "supply," "storm," "terrorism," and "disaster" have high inclusion probabilities, indicating they are significant predictors for the GSCPI. Specifically, higher mentions of "supply" and "storm" related lexicon in the news are associated with a higher index value. These lexicons include words such as "dearth," "deficit," "shortage," "shortfall," "winter," "wintertime," "overwinter," "blizzard," "rain," etc. (see Appendix A – Topics Lexicon) which could summarize some of the most important sources of supply chain disruptions.

Conversely, a lower share of “disaster” and “terrorism” related lexicon in the news correlates with a higher GSCPI, suggesting that less frequent mentions of words such as “terrorism,” “terrorist act,” “terrorist,” “calamity,” “cataclysm,” “catastrophe,” etc., correspond to an increased GSCPI. This counterintuitive finding could be related to the fact that during periods of heightened supply chain pressure, news coverage (related to supply chain disruptions) might shift focus towards specific supply chain issues, reducing the relative frequency of coverage on disasters and terrorism. When the media focuses more on immediate supply chain disruptions, the relative mention of broader disaster and terrorism topics decreases, even though the actual pressure on the supply chain, as measured by the GSCPI, is increasing. This shift in focus could lead to a lower share of disaster and terrorism-related lexicon in the news during times of high supply chain pressure.

The relevance of a variable as a predictor for the outcome is measured by its PIP and marginal effects. Figure 3 presents the marginal effects of the regressors on the GSCPI, representing the outcome’s response to a one-unit change in each variable. The figure shows that the sentiment score has the largest negative marginal effect on the outcome, underscoring the significance of this variable in the model. This implies that the GSCPI is highly responsive to changes in the news sentiment. Conversely, the figure indicates a strong positive marginal effect of “storm” on the outcome, confirming the relevance of the frequency of storm-related topics in the news with the GSCPI. This positive correlation can be explained by the fact that severe storms often cause significant disruptions in supply chains. These disruptions can range from delays in transportation and delivery to damage to infrastructure and goods, which can increase the GSCPI. Therefore, more frequent mentions of storms in the news reflect higher supply chain pressures, leading to an increase in the GSCPI.

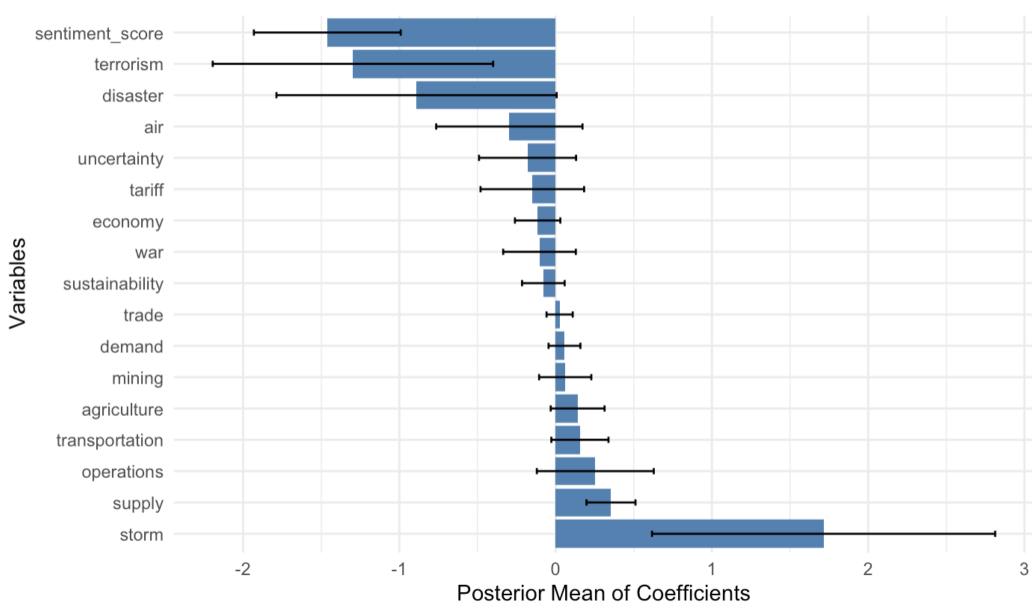


Figure 3. Marginal effects

The resulting model obtains the prediction densities for the entire dataset. Figure 4 shows the prediction densities for the 25th and 75th quantiles, revealing that the model successfully tracks significant supply chain disruption periods, such as the peaks experienced between 2020 and 2022 and the trough seen in 2023-2024. Equivalent Ordinary Least Squares (OLS) indicators demonstrate that the developed model captures about 80% of the variables influencing the GSCPI and is statistically significant at a 99% confidence level.

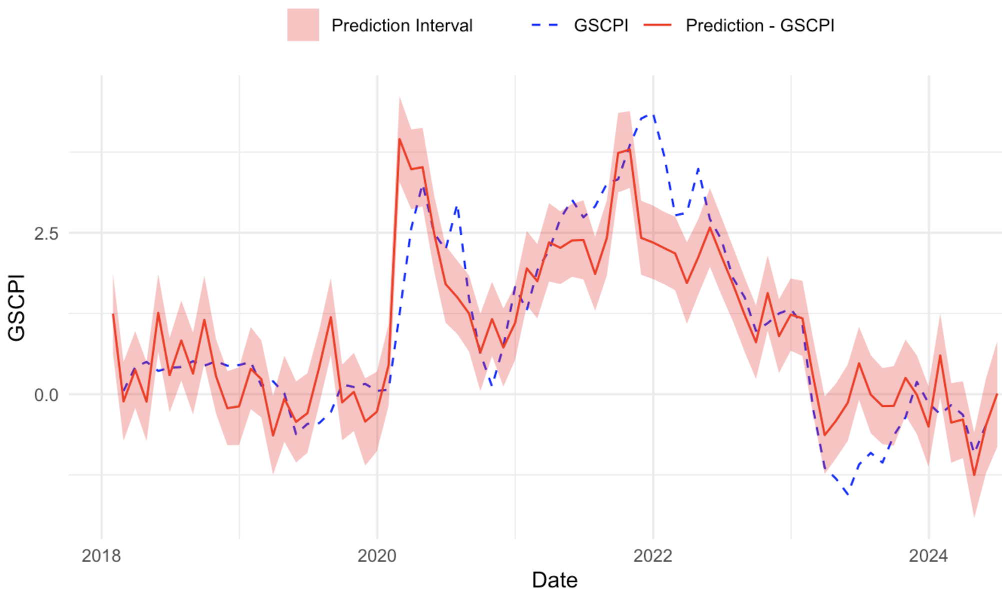


Figure 4. GSCPI vs fitted values.

Task 3 – Dashboard Creation, Testing, Demonstration, and Incorporation of Feedback

The dashboard serves as the visualization component of the tool, providing an interactive framework to present insights on news, metrics, and predictive models to facilitate decision-making based on supply chain information. The team developed the dashboard for this project using Shiny [58], an open-source package offering a web framework for building and deploying interactive web applications. Specifically, the dashboard was implemented in R using the *'shinydashboard'* package [71], chosen for its ability to streamline communication and data transfer across different tool modules.

The dashboard is structured into five main tabs: Summary of News, News Metrics, Macroeconomic and Supply Chain Metrics, Predictive Modeling, and California Indicators, each serving distinct purposes as described below:

Summary of News

The first tab of the dashboard offers general information about the publication of news articles. It begins by presenting three statistics on the number of news published in the last 48 hours, 7 days, and 30 days, based on the outcomes of the text analytics module. A time-series object is created based on the sentiment score data using the *'xts'* [72] package in R. The time component of the time series corresponds to the publishing date, while the dependent variable reflects the count of news published on that day. The time series are plotted using the *'dygraphs'* package [73], enabling interactive charting of time-series data.

Additionally, the sentiment score database is utilized to track the evolution of sentiment scores over time. Similarly, the topics database displays how the share of the eight defined topics and their respective subtopics change over time. Interactive widgets are integrated into this tab to enhance user experience. This section also displays the California Index to provide insights into the state's relevance in supply-chain-related news. A range selector allows users to select a specific time range and zoom in on the data.

Furthermore, a slider lets users set a rolling period, smoothing out the time series display by plotting the moving average over the specified period. For instance, if the rolling period is set to 7, the time series will depict the moving average of the last seven days. Figure 5 shows a visualization of this tab.

News Metrics

This tab presents key results of the Natural Language Processing (NLP) of news articles, including geographical location, word clouds, sentiment of words, and top 10 locations, as presented below:

- Geographical location: This chart displays a heatmap of geographies mentioned in the news, providing insight into regions focused on supply chain risks. The location is presented at three levels: world (countries), U.S. (States of the U.S.), and California (cities in California), as shown in Figure 6. Interactive maps are plotted using the leaflet package [74] in R, allowing users to explore geographical data dynamically. A widget for selecting the geographical level enhances user interactivity.

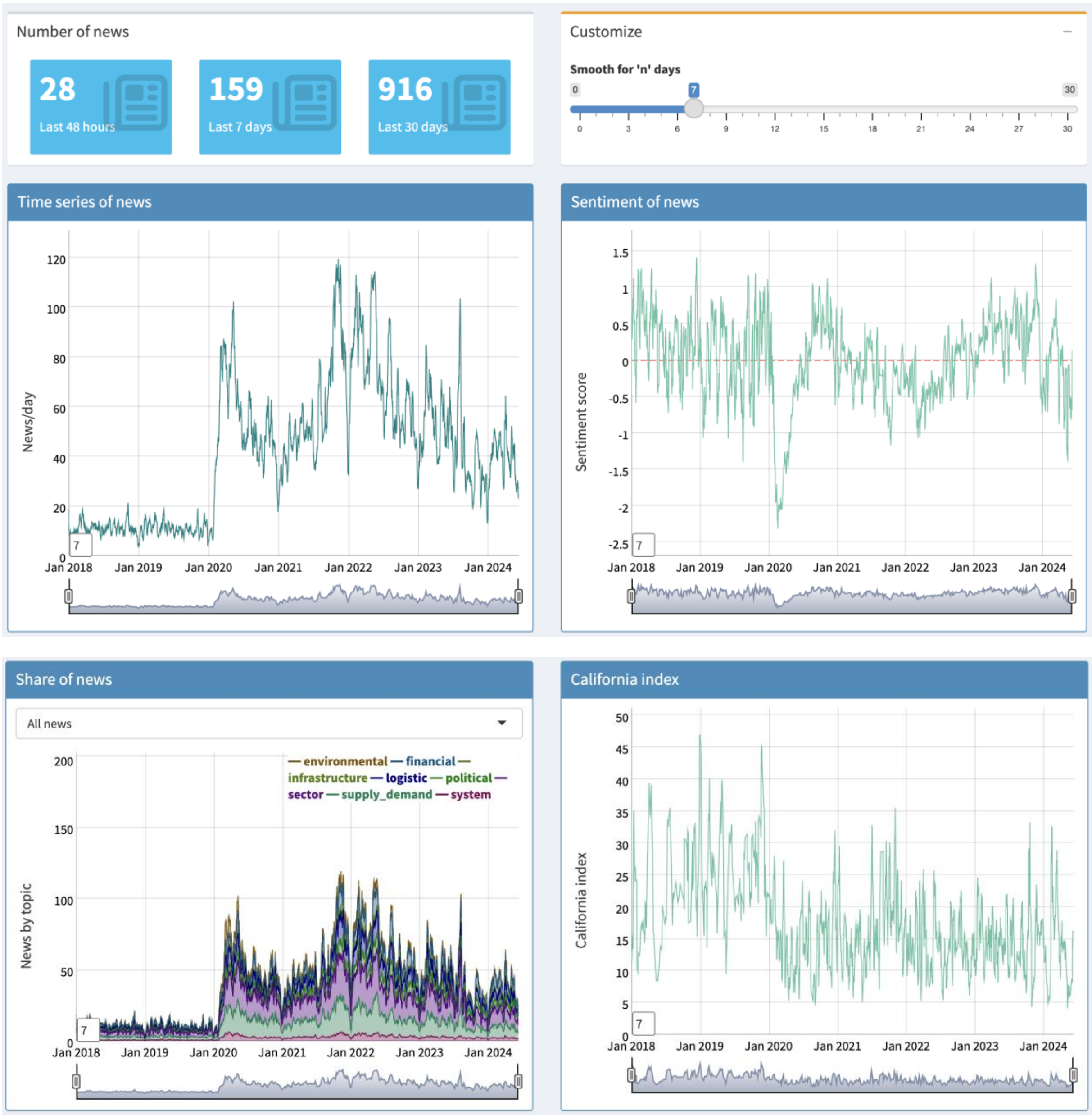
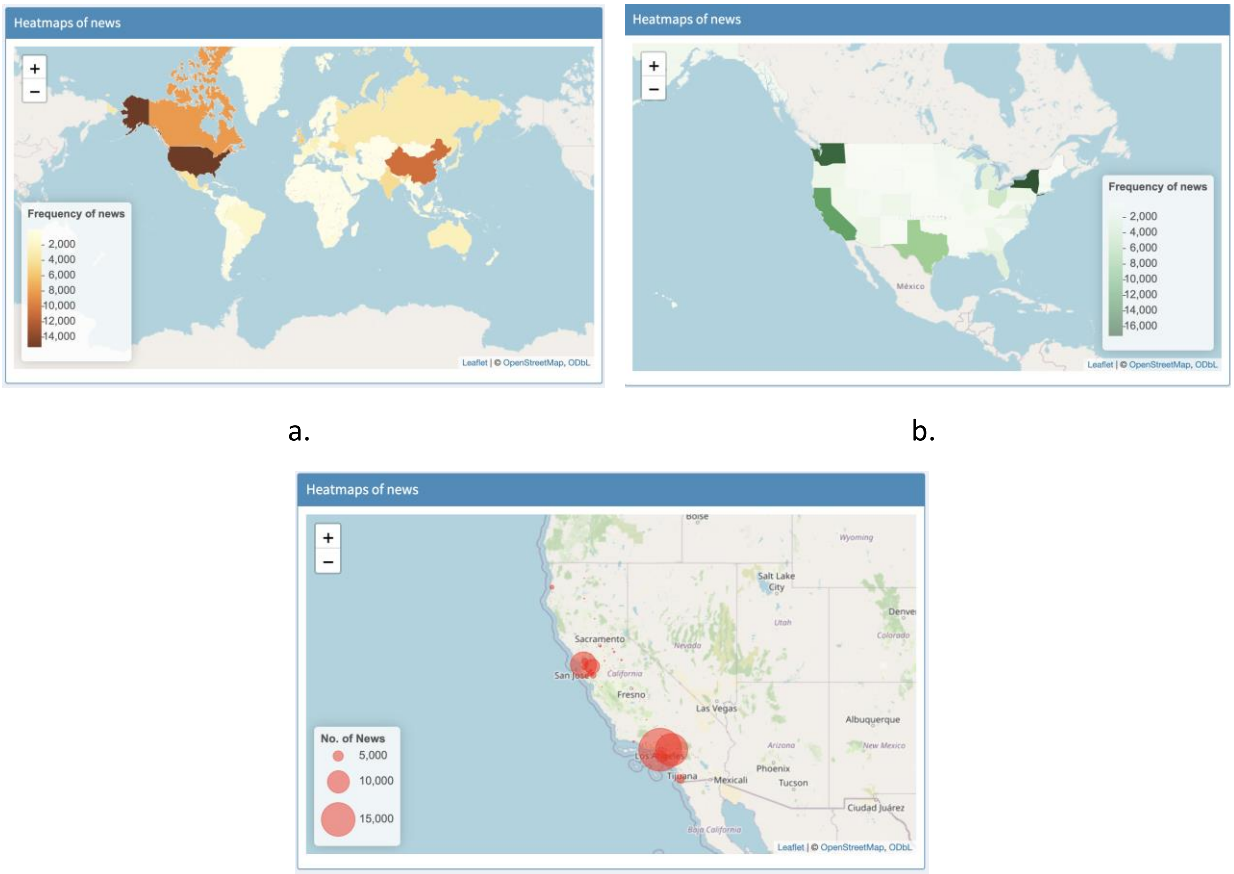


Figure 5. News time series tab.

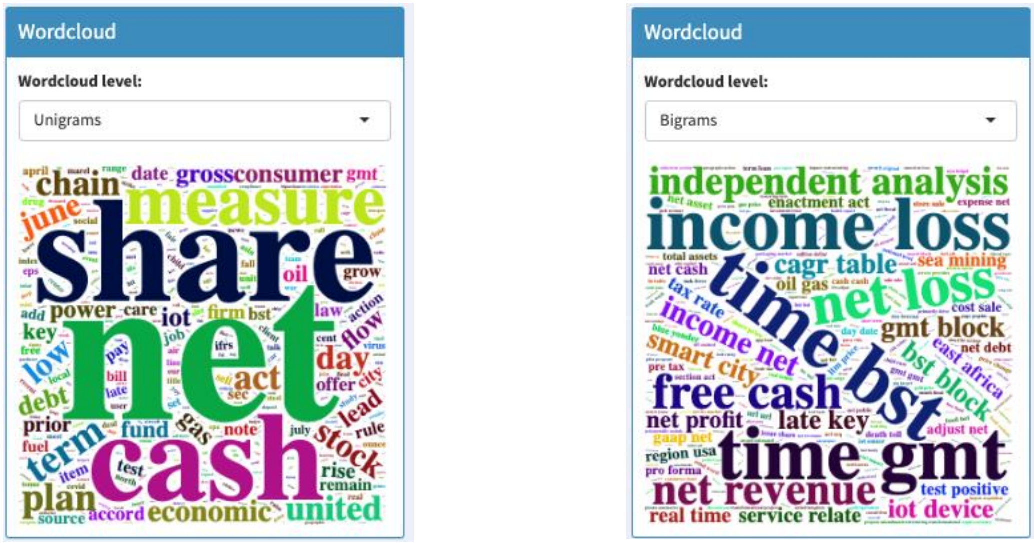


a. b.

c.

Figure 6. Geographical location of news, a. World, b. U.S., c. California.

- Wordcloud: N-grams are visually represented using word clouds through the *'wordclouds'* [75] package in r, highlighting unigrams and bigrams with the highest relevance (TF-IDF). Word clouds facilitate the identification of hot topics in the news, as depicted in Figure 7. Users can select the level of the word cloud (unigram or bigram) using an included widget.
- Words sentiments: A time series of the share of words with positive and negative sentiments is presented. The chart includes a rollover period for plotting the moving average. As shown in Figure 8, news presents a larger share of positive words. This is because the positive lexicon is more commonly used in news, and the 'AFINN' dictionary performs better in classifying positive and neutral words [54]. However, the trend in sentiment over time can indicate supply chain disruptions. For instance, a change in sentiment trends coinciding with the onset of the COVID-19 pandemic in January 2020 suggests potential disruptions.



a. b.

Figure 7. Word clouds, a. unigrams, b. bigrams.

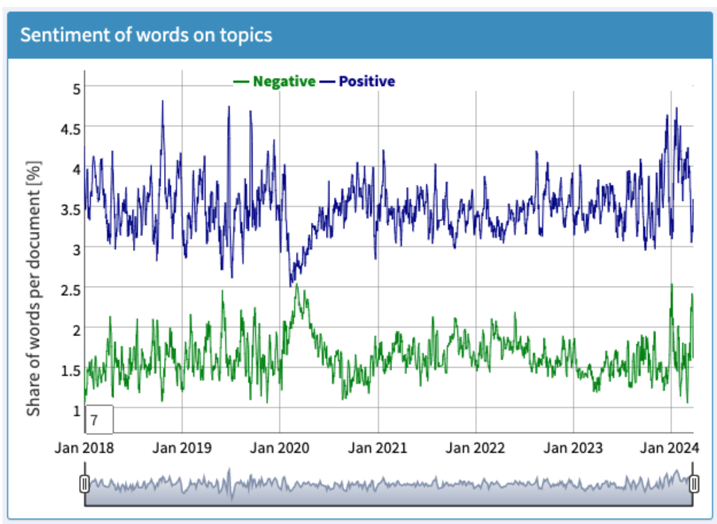


Figure 8. Sentiment of words chart.

- In addition, a table with the top 10 locations is provided in this tab. The user can define the time window for generating this table. Three options are provided: last 48 hours, last 7 days, and last 30 days. Furthermore, the main widgets are included in the dashboard. First, a date slider allows users to filter news within specific time windows, enabling temporal analysis of supply chain trends. Second, a topic selector that reproduces the charts based on news articles with a share of a selected topic exceeding 25%, providing focused insights into specific supply chain topics. Figure 9 provides a whole picture of this tab.

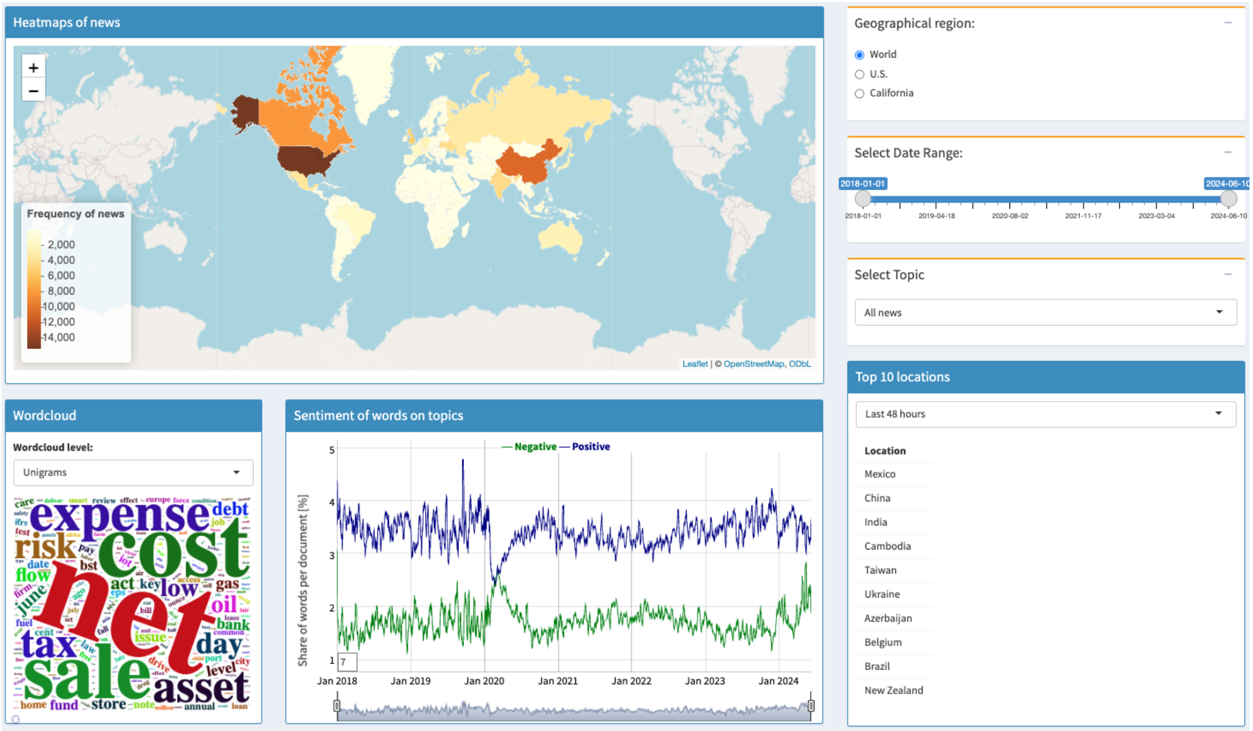


Figure 9. News Metrics tab.

See Appendix C – Empirical Validation in Task 2 for examples of empirical validation of the news data trends in the dashboard.

Macroeconomic and Supply Chain Metrics

The third tab of the dashboard visualizes macroeconomic metrics for the U.S. and California, the GSCPI, and data on seaport imports and exports at the national and state levels. Interactive time series charts are plotted based on three main widgets:

1. **Region selector:** Allows users to choose the area to analyze (U.S. or California), facilitating focused exploration of specific macroeconomic metrics.
2. **Metric selector:** Once a region is selected, the backend filters all relevant metrics, providing users with available indicators corresponding to the chosen region. This enables users to tailor the analysis to their specific interests.
3. **Seasonal adjustment:** Some indicators offer two versions – seasonally and non-seasonally adjusted. A selector is provided to plot the desired chart, allowing users to compare trends with and without seasonal fluctuations.

Additionally, the dashboard describes the selected metric to provide users context and better inform their analysis. This descriptive feature enhances user understanding and interpretation of the visualized data. This tab is shown in Figure 10.

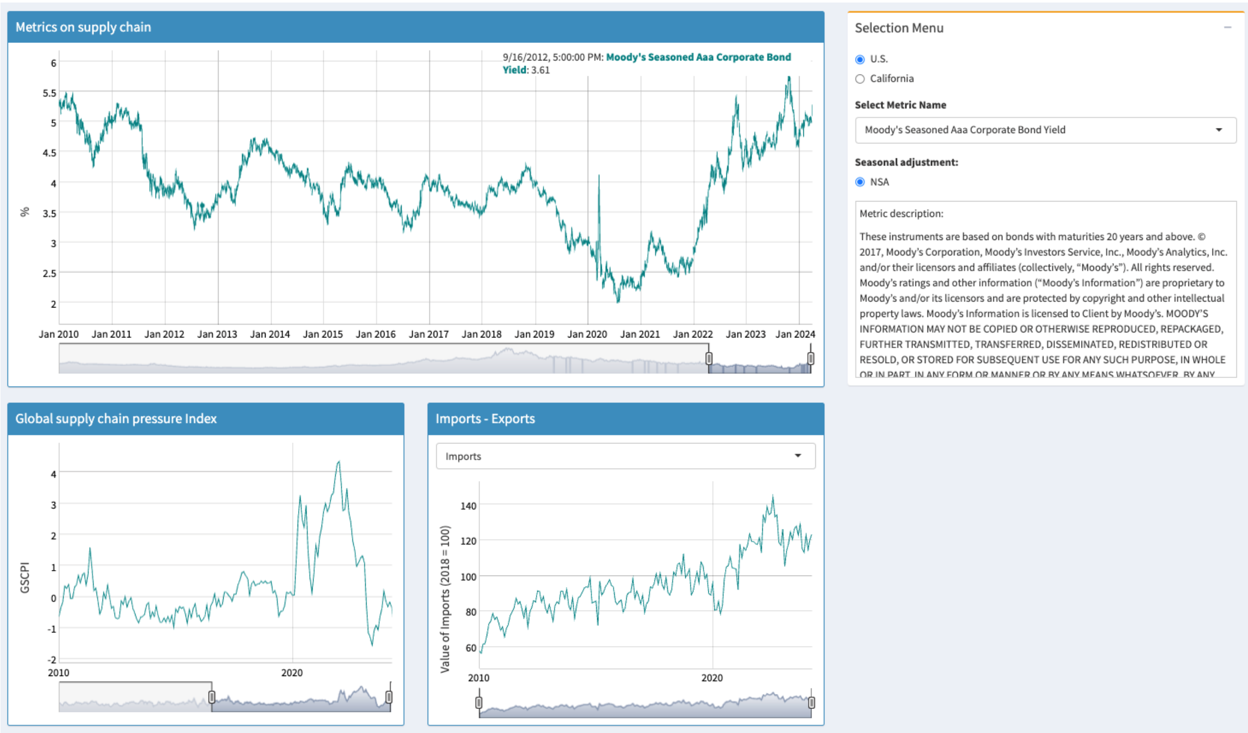


Figure 10. Macroeconomic and supply Chain chart.

Predictive Modeling

The fourth tab of the interface showcases the predictive models developed for relevant macroeconomic and supply chains presented in tab 3, such as the Consumer Price Index, Non-farm Employment, GSCPI, and imports and exports. Some of these indicators are modeled at the California and U.S. levels. Below is a breakdown of the main components of this section:

1. **Prediction chart:** This interactive plot displays the behavior of the actual metric and the predicted values obtained through the BMA model. This chart is plotted using *'dygraphs'* so users can easily identify the latest published value and forecasted values for future months.
2. **Model inclusion:** This chart illustrates the participation of different variables in predicting the selected outcomes. It depicts the inclusion of the variables selected in the forecasting models after evaluating lag and filtering out the regressors below the minimum PIP.
3. **Marginal effects:** Using the *'plotly'* package [76], this figure presents Posterior Mean Coefficient intervals for selected variables used in modeling outcomes. Its interactive features enhance usability.

Additionally, this tab provides OLS-equivalent statistics such as R2, Error Sum of Squares (ESS), Residual Sum of Squares (RSS), Total Sum of Squares (TSS), and F-Stat. A widget is also included to enable users to select the model they wish to explore further. Figure 11 offers a snapshot of this tab's interface.

Figure 11. Predictive modeling tab.

California Indicators

The last tab of the interface provides essential data concerning key indicators specific to the state of California. This includes information on the two prominent seaports within the state, L.A. and L.B., along with highway data. This section offers insights into the average time at berth for the Port of L.A., which indicates dwell time and reflects container terminal efficiency [77]. Additionally, it presents data on truck turn times, encompassing average queue time, terminal time, and total time. These metrics are instrumental in gauging the extent of congestion at the port. Figure 12 depicts the layout of this tab for reference.

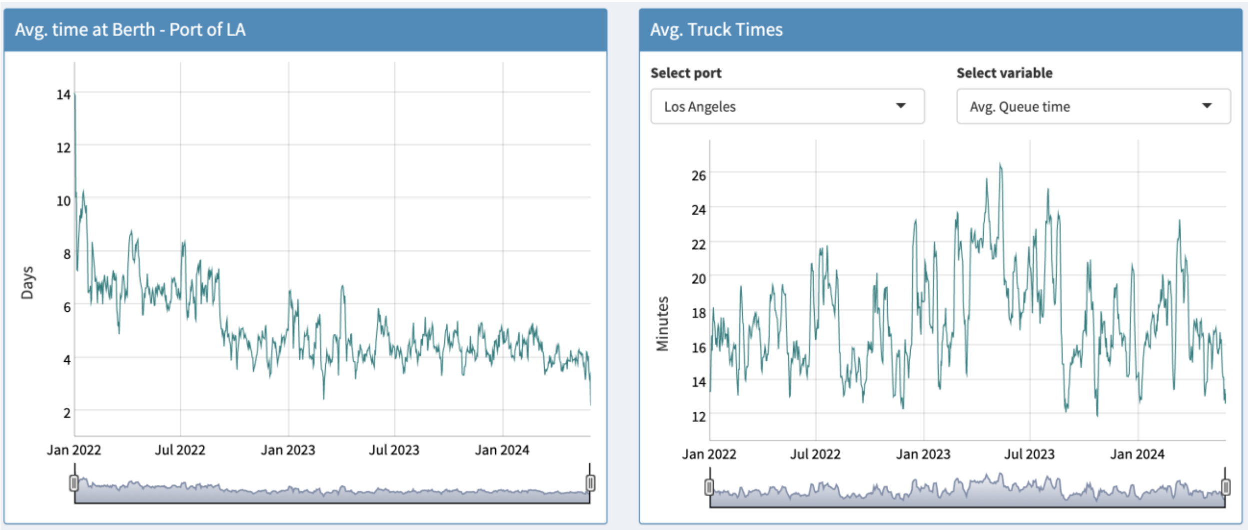


Figure 12. California indicators tab.

Conclusions

Text-mining and Bayesian Model Averaging (BMA) techniques offer a promising approach for forecasting the impacts of supply chain disruptions on key logistics and macroeconomic indicators such as the Consumer Price Index (CPI), employment, the Global Supply Chain Pressure Index (GSCPI), and seaport imports and exports.

The empirical results demonstrate that using BMA at the subtopic level enhances the model's granularity, improving the predictive accuracy. However, there is a need to evaluate the tradeoffs between model complexity and predictive performance. Filtering regressors based on their Posterior Inclusion Probabilities (PIPs) effectively achieved this tradeoff.

Sentiment analysis has emerged as one of the most critical factors in predicting economic outcomes. The sentiment score exhibited a significant negative correlation with the GSCPI, indicating that more negative news sentiment correlates with higher supply chain pressures. This relationship underscores the importance of the tone and content of news coverage in shaping perceptions and expectations about supply chain performance. As such, tracking sentiment trends can serve as an early indicator of potential stress within the supply chain, providing actionable insights for preemptive decision-making.

Our model effectively captures the compounded effects of simultaneous disruptions, but it is limited in isolating the influence of individual events due to supply chains' interconnected and global nature. Nevertheless, the tool's ability to prioritize and use significant disruptions to forecast their impact on the global supply chain, as indicated by the GSCPI and other relevant indicators, provides valuable insight. This capacity to analyze multiple concurrent events is crucial for formulating resilient supply chain management strategies.

This tool enables early detection of potential supply chain risks and disruptions, allowing decision-makers to respond with more agility and data-driven strategies. Future work will integrate sustainability indicators to assess the impact of disruptions on energy consumption and CO_2 emissions. By including these metrics, the tool will support strategies that address immediate operational risks while promoting long-term sustainability and resilience. These insights could inform the development of stronger policies that enhance the adaptability and sustainability of global supply chains in response to both anticipated and unexpected challenges.

References

[1] Katsaliaki K, Galetsi · P, Kumar · S, Kumar S, Galetsi P. Supply chain disruptions and resilience: a major review and future research agenda 2022;319:965–1002. https://doi.org/10.1007/s10479-020-03912-1.

[2] United States Trade Representative. Building Resilient and Secure Supply Chains Through Trade 2022. https://ustr.gov/about-us/policy-offices/press-office/blogs-and-op-eds/2022/april/building-resilient-and-secure-supply-chains-through-trade (accessed March 18, 2024).

[3] The White House. Issue Brief: Supply Chain Resilience 2023. https://www.whitehouse.gov/cea/written-materials/2023/11/30/issue-brief-supply-chain-resilience/ (accessed March 18, 2024).

[4] The White House. Executive Order on America's Supply Chains 2021. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/02/24/executive-order-on-americas-supply-chains/ (accessed March 18, 2024).

[5] U.S. DOT. Supply Chain Assessment of the Transportation Industrial Base: Freight and Logistics. 2022.

[6] U.S. DOT. Strategic Plan F.Y. 2022-2026. 2022.

[7] Rivera-Royero D, Jaller M, Jenn A. Impacts of Precautionary and Opportunistic Buying Behaviors and Supply Issues on Supply Chain Resilience During the COVID-19 Pandemic. Transportation Research Record: Journal of the Transportation Research Board 2022. https://doi.org/10.1177/03611981221124880.

[8] Delgado C, Castelo BM. Supply Chain Management. Encyclopedia of Corporate Social Responsibility 2013:2349–57. https://doi.org/10.1007/978-3-642-28036-8_139.

[9] Delgado M, Mills KG. The Supply Chain Economy: A New Framework for Understanding Innovation and Services. 2017.

[10] Delgado M, Mills KG. The supply chain economy: A new industry categorization for understanding innovation in services. Res Policy 2020;49:104039. https://doi.org/10.1016/J.RESPOL.2020.104039.

[11] Randstad. Cause and effect of supply chain disruptions 2022. https://www.randstad.com/workforce-insights/workforce-management/cause-effect-supply-chain-disruptions/ (accessed March 18, 2024).

[12] Kalemli-Özcan Julian di Giovanni Álvaro Silva Muhammed Yıldırım Ş. Global supply chain pressures, international trade and inflation Challenges for monetary policy in a rapidly changing world. 2022.

[13] Santacreu AM, LaBelle J. Supply Chain Disruptions and Inflation During COVID-19. vol. 2022. Federal Reserve Bank of St. Louis; 2022. https://doi.org/10.20955/ES.2022.11.

[14] McKinsey. Risk, resilience, and rebalancing in global value chains. 2020.

[15] Xu S, Zhang X, Feng L, Yang W. Disruption risks in supply chain management: a literature review based on bibliometric analysis. Int J Prod Res 2020;58:3508–26. https://doi.org/10.1080/00207543.2020.1717011.

[16] Carissimi MC, Creazza A, Colicchia C. Crossing the chasm: investigating the relationship between sustainability and resilience in supply chain management. Cleaner Logistics and Supply Chain 2023;7:100098. https://doi.org/10.1016/J.CLSCN.2023.100098.

[17] Saidi D, Ait Bassou A, Alami J El, Hlyal M, Rabat I. Sustainability and Resilience Analysis in Supply Chain Considering Pricing Policies and Government Economic Measures. IJACSA) International Journal of Advanced Computer Science and Applications 2024;15.

[18] Han N, Um J. Risk management strategy for supply chain sustainability and resilience capability. Risk Management 2024;26:1–26. https://doi.org/10.1057/S41283-023-00138-W/TABLES/6.

[19] Pahwa A, Jaller M. Assessing last-mile distribution resilience under demand disruptions. Transp Res E Logist Transp Rev 2023;172:103066. https://doi.org/10.1016/J.TRE.2023.103066.

[20] Tukamuhabwa BR, Stevenson M, Busby J, Zorzini M. Supply chain resilience: definition, review and theoretical foundations for further study. Int J Prod Res 2015;53:5592–623. https://doi.org/10.1080/00207543.2015.1037934.

[21] U.S. DOT. Research, Development, and Technology Strategic Plan Fiscal Year 2022-2026. 2022.

[22] United States Department of State. Joint Statement on Cooperation on Global Supply Chains 2022. https://www.state.gov/supply-chain-ministerial-joint-statement/ (accessed March 18, 2024).

[23] The White House. FACT SHEET: President Biden Announces New Actions to Strengthen America's Supply Chains, Lower Costs for Families, and Secure Key Sectors 2023. https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/27/fact-sheet-president-biden-announces-new-actions-to-strengthen-americas-supply-chains-lower-costs-for-families-and-secure-key-sectors/ (accessed February 17, 2024).

[24] Ingle C, Bakliwal D, Jain J, Singh P, Kale P, Chhajed V. Demand Forecasting: Literature Review On Various Methodologies. 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021 2021. https://doi.org/10.1109/ICCCNT51525.2021.9580139.

[25] Barbaglia L, Consoli S, Manzan S. Forecasting with Economic News. Journal of Business & Economic Statistics 2023;41:708–19. https://doi.org/10.1080/07350015.2022.2060988.

[26] Joshi S, Sharma M, Rathi S. Forecasting in service supply chain systems: A state-of-the-art review using latent semantic analysis. Advances in Business and Management Forecasting 2017;12:181–212. https://doi.org/10.1108/S1477-407020170000012011/FULL/XML.

[27] Öztayşi B, Bolturk E. Fuzzy methods for demand forecasting in supply chain management. Studies in Fuzziness and Soft Computing 2014;313:243–68. https://doi.org/10.1007/978-3-642-53939-8_11/TABLES/3.

[28] Barbaglia L, Consoli S, Manzan S. Forecasting GDP in Europe with textual data 2024. https://doi.org/10.1002/jae.3027.

[29] Feuerriegel S, Gordon J. News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions. Eur J Oper Res 2019;272:162–75. https://doi.org/10.1016/J.EJOR.2018.05.068.

[30] Aprigliano V, Emiliozzi S, Guaitoli G, Luciani A, Marcucci J, Monteforte L. The power of text-based indicators in forecasting Italian economic activity. Int J Forecast 2023;39:791–808. https://doi.org/10.1016/J.IJFORECAST.2022.02.006.

[31] Steel M. Bayesian Model Averaging and Forecasting 2011.

[32] Kumar Sharma S, Sharma S. Developing a Bayesian Network Model for Supply Chain Risk Assessment. Supply Chain Forum: An International Journal 2015;16:50–72. https://doi.org/10.1080/16258312.2015.11728693.

[33] Bencivelli L, Marcellino M, Moretti G. Forecasting economic activity by Bayesian bridge model averaging. Empir Econ 2017;53:21–40. https://doi.org/10.1007/S00181-016-1199-9/TABLES/7.

[34] Vosseler A, Weber E. Forecasting seasonal time series data: a Bayesian model averaging approach. Comput Stat 2018;33:1733–65. https://doi.org/10.1007/S00180-018-0801-3/FIGURES/9.

[35] Kapetanios G, Labhard V, Price S. Forecasting Using Bayesian and Information-Theoretic Model Averaging. Journal of Business & Economic Statistics 2008;26:33–41. https://doi.org/10.1198/073500107000000232.

[36] Chatterjee S, Ranjan Chaudhuri ·, Vrontis D, Chaudhuri R. Role of fake news and misinformation in supply chain disruption: impact of technology competency as moderator. Ann Oper Res 2023;327:659–82. https://doi.org/10.1007/s10479-022-05001-x.

[37] Nexis Uni. Academic Research Tool for Universities & Libraries 2024. https://www.lexisnexis.com/en-us/professional/academic/nexis-uni.page (accessed March 15, 2024).

[38] Shapiro AH, Sudhof M, Wilson DJ. Measuring news sentiment. J Econom 2022;228:221–43. https://doi.org/10.1016/J.JECONOM.2020.07.053.

[39] Python.org. Python 2024. https://www.python.org/ (accessed March 27, 2024).

[40] NewsAPI. Documentation - News API 2024. https://newsapi.org/docs (accessed March 27, 2024).

[41] Schulze L. Newsanchor 2019. https://cran.r-project.org/web/packages/newsanchor/vignettes/usage-newsanchor.html (accessed March 27, 2024).

[42] R. The Comprehensive R Archive Network 2024. https://cran.rstudio.com/ (accessed March 27, 2024).

[43] Wickham H. httr: Tools for Working with URLs and HTTP 2023. https://CRAN.R-project.org/package=httr (accessed March 27, 2024).

[44] Federal Reserve Bank of St. Louis. What is FRED? 2024. https://fredhelp.stlouisfed.org/fred/about/about-fred/what-is-fred/ (accessed March 27, 2024).

[45] Boysel S. Getting started with fredr 2021. https://cran.r-project.org/web/packages/fredr/vignettes/fredr.html (accessed March 27, 2024).

[46] Federal Reserve Bank of New York. Global Supply Chain Pressure Index (GSCPI) 2022. https://www.newyorkfed.org/research/policy/gscpi#/overview (accessed April 29, 2024).

[47] U.S. Census Bureau. International Trade Data Main Page 2024.

[48] U.S. Census Bureau. International Trade Data API User Guide. 2024.

[49] Wabtec. Port Optimizer - Control Tower 2024. https://signal.portoptimizer.com/ (accessed March 27, 2024).

[50] CHP. CHP Headquarters (@CHP_HQ) / X 2024. https://twitter.com/CHP_HQ?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor (accessed March 27, 2024).

[51] Chu CY, Park K, Kremer GE. A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. Advanced Engineering Informatics 2020;45:101053. https://doi.org/10.1016/J.AEI.2020.101053.

[52] Nandwani P, Verma R. A review on sentiment analysis and emotion detection from text 2021;11:81. https://doi.org/10.1007/s13278-021-00776-6.

[53] Onyenwe I, Nwagbo S, Mbeledogu N, Onyedinma E. The impact of political party/candidate on the election results from a sentiment analysis perspective using #AnambraDecides2017 tweets. Soc Netw Anal Min 2020;10:1–17. https://doi.org/10.1007/S13278-020-00667-2/FIGURES/14.

[54] Nkongolo Wa Nkongolo M. News Classification and Categorization with Smart Function Sentiment Analysis. International Journal of Intelligent Systems 2023;2023. https://doi.org/10.1155/2023/1784394.

[55] Kaur A, Gupta V. A Survey on Sentiment Analysis and Opinion Mining Techniques 2013. https://doi.org/10.4304/jetwi.5.4.367-371.

[56] Petersens R. AFINN 2011. https://www2.imm.dtu.dk/pubdb/pubs/6010-full.html (accessed March 27, 2024).

[57] Muller K, Wickham H, James DA, Falcon S. SQLite Interface for R 2024. https://CRAN.R-project.org/package=RSQLite (accessed March 27, 2024).

[58] SpaCy. spaCy: Industrial-strength Natural Language Processing in Python 2024. https://spacy.io/ (accessed March 27, 2024).

[59] Schmitt X, Kubler S, Robert J, Papadakis M, Letraon Y. A Replicable Comparison Study of NER Software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. 2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019 2019:338–43. https://doi.org/10.1109/SNAMS.2019.8931850.

[60] Piantadosi ST. Zipf's word frequency law in natural language: A critical review and future directions 2014. https://doi.org/10.3758/s13423-014-0585-6.

[61] Zhang Y, Zhou Y, Yao JT. Feature Extraction with TF-IDF and Game-Theoretic Shadowed Sets. Communications in Computer and Information Science 2020;1237 CCIS:722–33. https://doi.org/10.1007/978-3-030-50146-4_53/TABLES/4.

[62] Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, et al. Quanteda: Quantitative Analysis of Textual Data 2023. https://CRAN.R-project.org/package=quanteda (accessed March 27, 2024).

[63] Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, et al. quanteda: An R package for the quantitative analysis of textual data. The Journal of Open Source Software 2018. https://doi.org/10.21105/joss.00774.

[64] Watanabe K. R: Semi-supervised Bayesian multinomial model for geographical... 2018. https://search.r-project.org/CRAN/refmans/newsmap/html/textmodel_newsmap.html (accessed March 27, 2024).

[65] Watanabe K. Newsmap. Digital Journalism 2018;6:294–309. https://doi.org/10.1080/21670811.2017.1293487.

[66] Gabrielyan D, Masso J, Uusküla L. Mining News Data for the Measurement and Prediction of Inflation Expectations 2020:253–71. https://doi.org/10.1007/978-3-030-56219-9_17.

[67] Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Mon Weather Rev 2005;133:1155–74. https://doi.org/10.1175/MWR2906.1.

[68] Sloughter JML, Raftery AE, Gneiting T, Fraley C. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. Mon Weather Rev 2007;135:3209–20. https://doi.org/10.1175/MWR3441.1.

[69] Feldkircher M, Zeugner S. R package BMS - Bayesian Model Averaging 2024. http://bms.zeugner.eu/ (accessed June 10, 2024).

[70] Hinne M, Gronau QF, van den Bergh D, Wagenmakers EJ. A Conceptual Introduction to Bayesian Model Averaging. Adv Methods Pract Psychol Sci 2020;3:200–15. https://doi.org/10.1177/2515245919898657/ASSET/IMAGES/LARGE/10.1177_2515245919898657-FIG4.JPEG.

[71] Chang W, Borges Ribeiro B. Create Dashboards with “Shiny” [R package shinydashboard version 0.7.2] 2021. https://CRAN.R-project.org/package=shinydashboard (accessed March 27, 2024).

[72] Ryan J, Ulrich J, Bennet R, Joy C. eXtensible Time Series 2024. https://CRAN.R-project.org/package=xts (accessed March 27, 2024).

[73] dygraphs. dygraphs for R 2024. https://rstudio.github.io/dygraphs/ (accessed March 27, 2024).

[74] Cheng J, Karambelkar B, Xie T. An R Interface to Leaflet Maps 2023. https://rstudio.github.io/leaflet/ (accessed March 27, 2024).

[75] Fellows I. wordcloud: Word Clouds 2018. https://cran.r-project.org/web/packages/wordcloud/ (accessed March 27, 2024).

[76] Plotly. Plotly r graphing library in R 2024. https://plotly.com/r/ (accessed June 10, 2024).

[77] Notteboom T, Pallis A, Rodrigue J-P. Port Economics, Management and Policy. Port Economics, Management and Policy 2021. https://doi.org/10.4324/9780429318184.

Appendix A – Topics Lexicon

Topic	Subtopic	Words
political	terrorism	act of terrorism, terrorism, terrorist act, terrorist, terror, threat, affright, brat, holy terror, little terror, panic, scourge, menace
	war	state of war, war, warfare, conflict, battle, hostility, attack, hostile, combat, dispute, fight, contravene, aggression, enmity, blast, onslaught, plan of attack, tone-beginning, aggress, assail, assault, lash out, snipe, uncongenial, unfriendly, armed combat, fighting
	tariff	revenue enhancement, tax, taxation, tariff, charge per unit
	regulation	regulation, ordinance, regularisation, regularization, regulating, rule, regulatory, policy, bureau, bureaucracy, control, govern, law, authority, federal agency, government agency, office, bureaucratism, command, harness, regularise, regularize, regulate, jurisprudence
	sustainability	decarbonization, emissions, sustainable, sustainability, zero-emissions, clean, carbon, co2, carbon dioxide, pollution, pollutant, tailpipe, carbon emissions, waste, conservation, climate change, global warming, climate, temperature, footprint, greenhouse, greenhouse gas, atmosphere, biofuel, ozone, methane, ice loss, acidification, particulate matter, renewable, environment, environmental, uncontaminating, contamination, waste material, waste product, preservation, global climate change, clime, greenhouse emission, particulate, environs
environmental	storm	winter, wintertime, overwinter, blizzard, rain, storm, hurricane, cyclone, cloudburst, rash, snowstorm, pelting, rainfall, rainwater, rain down, tempest, violent storm, downpour, waterspout
	flood	alluvion, deluge, flood, flood tide, flowage, inundation, outpouring, overflow, rising tide, torrent, inundate, stream, tide, sealevel, violent stream, current, flow
	earth	earthquake, quake, seism, temblor, seismic, tremor, tsunami, volcano, volcanic, landslide, avalanche, mudflow, rockslide, rockfall, seismal, earth tremor, microseism, landslip
	drought	drought, drouth, heatwave, heat, high temperature, hotness, warmth
	disaster	calamity, cataclysm, catastrophe, disaster, tragedy, weather, natural, nuclear, upwind, weather(a), atmospheric condition, weather condition
	pandemic	pandemic, coronavirus, epidemic, virus, disease, illness, covid
financial	economy	fluctuation, variation, wavering, inflation, price, gdp, trade, turbulence, downturn, deflation, rising prices, cost, monetary value, gross domestic product, trade in, turbulency, downswing
	finance	loan, loanword, budget, profit, credit, bankruptcy, bankrupt, asset, capital, earnings, gain, lucre, net, net income, net profit, profits, benefit, turn a profit, deferred payment, failure, insolvent, washington
supply_demand	supply	dearth, deficit, famine, shortage, shortfall, provision, undersupply, scarcity, overstock, surplus, overproduction, supplier, production, manufacturing, import, supply, supplying, scarceness, excess, redundant, spare, supererogatory, superfluous, nimiety, surplusage, overrun, provider, output, product, importation, importee, provide
	demand	retail merchant, retailer, customer, export, store, market, retail, inventory, consumption, purchase, demand, sale, client, exportation, depot, shop, food market, grocery, grocery store, inventorying, stock-taking, stock list, stocktaking, stock-take, take stock, economic consumption, expenditure, ingestion, intake, phthisis, pulmonary tuberculosis, uptake, buy
logistic	facility	facility, facilities, warehouse, distribution center, infrastructure, storage warehouse
	operations	bullwhip, distribution, delay, storage, strike, dwell, capacity, efficacy, safety, dissatisfaction, congestion, utilization, hold, wait, storehouse, warehousing, work stoppage, safe, safety device, usage, use, utilisation
	labor	labor, labour, labour party, drive, drudge, worker, trucker, driver, union, salary, wage, wages, employee, employment, employer, retention, truck driver, labor

		union, pay, remuneration, engage, payoff, reward, employ, engagement, exercise, work, holding
	transportation	fare, shipping, transport, transportation, trucking, rate, spot rate, bottleneck, ship, container, containers, containership, hauling, truckage, cargo ships, merchant marine, merchant vessels, embark, container ship, container vessel
system	uncertainty	doubt, doubtfulness, dubiety, dubiousness, incertitude, precariousness, uncertainness, uncertainty, disruption, closure, havoc, competitor, break, commotion, disturbance, interruption, perturbation, blockage, closedown, closing, shutdown, stop, stoppage, competition
	resilience	transparence, transparency, transparentness, trust, resilience, recover, response, responsiveness, adapt, adaptation, partnership, network, diverse, diversion, redundancy, redundant, flexibility, flexible, connection, connect, repair, corporate trust, reliance, trustfulness, trustingness, resiliency, recuperate, reactivity, adaption, adjustment, redundance, elastic, flexile, association, connectedness, connective, connector, connexion, associate, fix, fixing, reparation, restore
infrastructure	water	port, seaport, waterway, barge, shipbuilding, dock, dockage, dockyard, harbor, vessel, harbour, haven, watercourse, flatboat, hoy, docking facility, loading dock, pier, wharf, wharfage, docking, docking fee, moorage, vas, watercraft
	rail	rail, railing, rails, track, rail in, rail off, train, railway, railroad, locomotive, intermodal, yards, railport, railroad line, railroad track, railway line, railway system, railroad train, wagon train, locomotor, railway locomotive
	road	highway, main road, road, interstate, street, expressway, lane, truck, long-haul, parking, inland, route, fuel, diesel, fueling, fleet, interstate highway, freeway, motorway, state highway, superhighway, refueling
	air	aerodrome, airdrome, airport, aircraft, drone, airline, aeroplane, air cargo, droning, airway, airplane, plane
sector	agriculture	crop, harvest, cultivate, grain, agriculture, produce, fruit, vegetable, food, cereal, plantation, livestock, farming, timber, fish, farm, ranch, forestry, caryopsis, food grain, veggie, agricultural, land, forest, lumber, harvesting, reap, cattle farm, cattle ranch
	construction	building, construction, structure, concrete, home, residence, housing, rent, mortgage, family, house, household, residency, edifice, lease
	manufacturing	industry, manufacture, assembling, machine, automation, materials, automotive, textile, leather, wood, plastic, rubber, metal, machinery, computer, electronics, electronic, equipment, appliance, component, furniture, fabrication, fabricate, auto, automobile, car, motorcar, cloth, fabric, material, metallic, alloy
	mining	gas, coal, mine, drilling, exploration, crude, well, shale, oil sands, hydrocarbon, mining, mineral, ore, quarrying, fracking, gasolene, gasoline, natural gas, petrol, ember, oil production, geographic expedition, blunt, raw, unrefined, crude oil, fossil oil, petroleum, rock oil
	trade	retail, merchandise, services, service, sale, goods, store, non-store, commerce, e-commerce, vendors, dealer, health care, personal care, gas stations, gas station, gasoline station, gasoline stations, book, miscellaneous, wholesale, wholesaling, resale, warehouse, merchant, brokers, agents, broker, agent, market, grocery, ware, inspection and repair, overhaul, sales agreement, sales event, shop, stock, storehouse, commercialism, mercantilism, bargainer, trader, healthcare, filling station, petrol station, storage warehouse, merchandiser, food market, grocery store, marketplace, securities industry, commercialice, commercialize, foodstuff

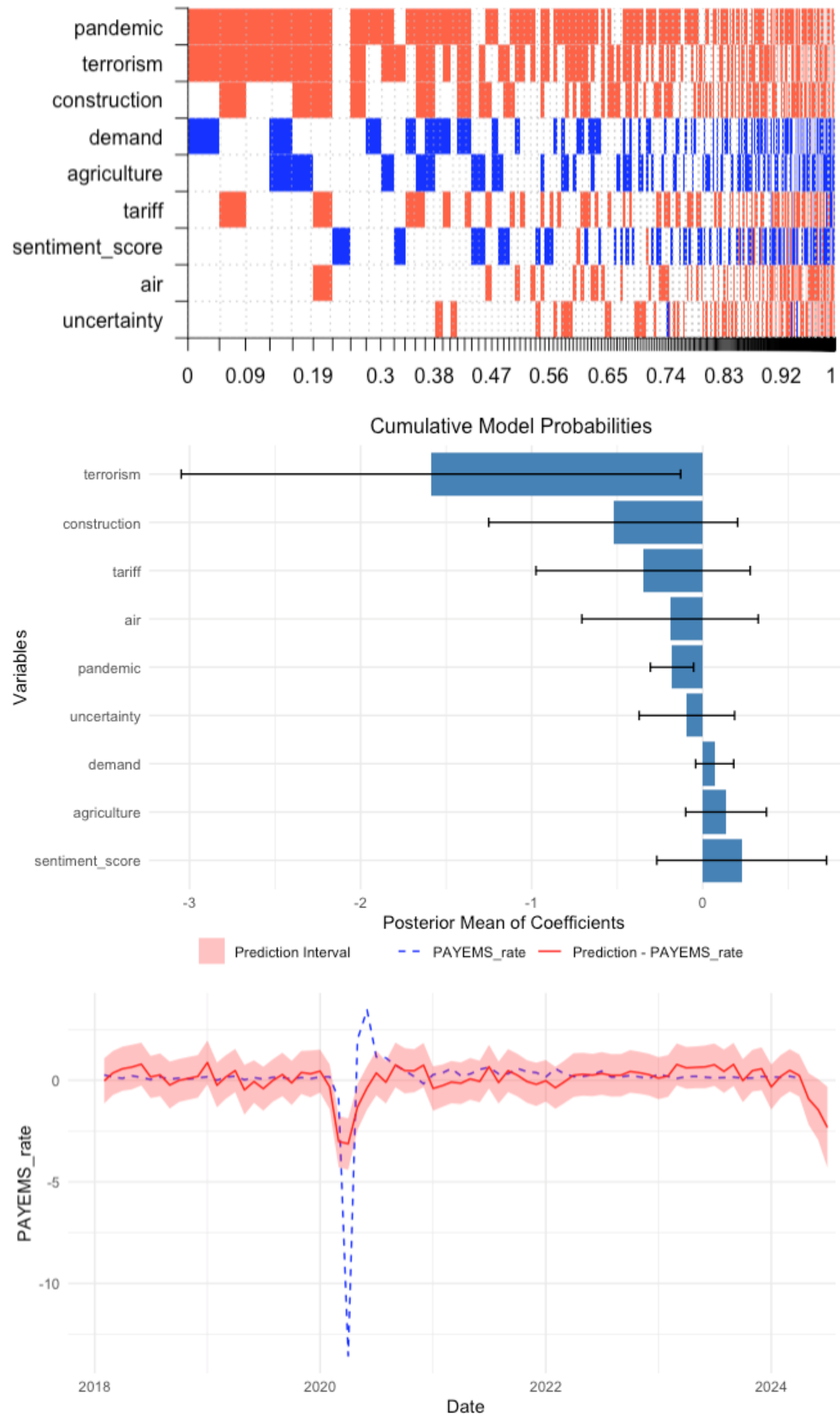
Appendix B – BMA Models

Non-farm Employment – U.S.

a)

b)

c)

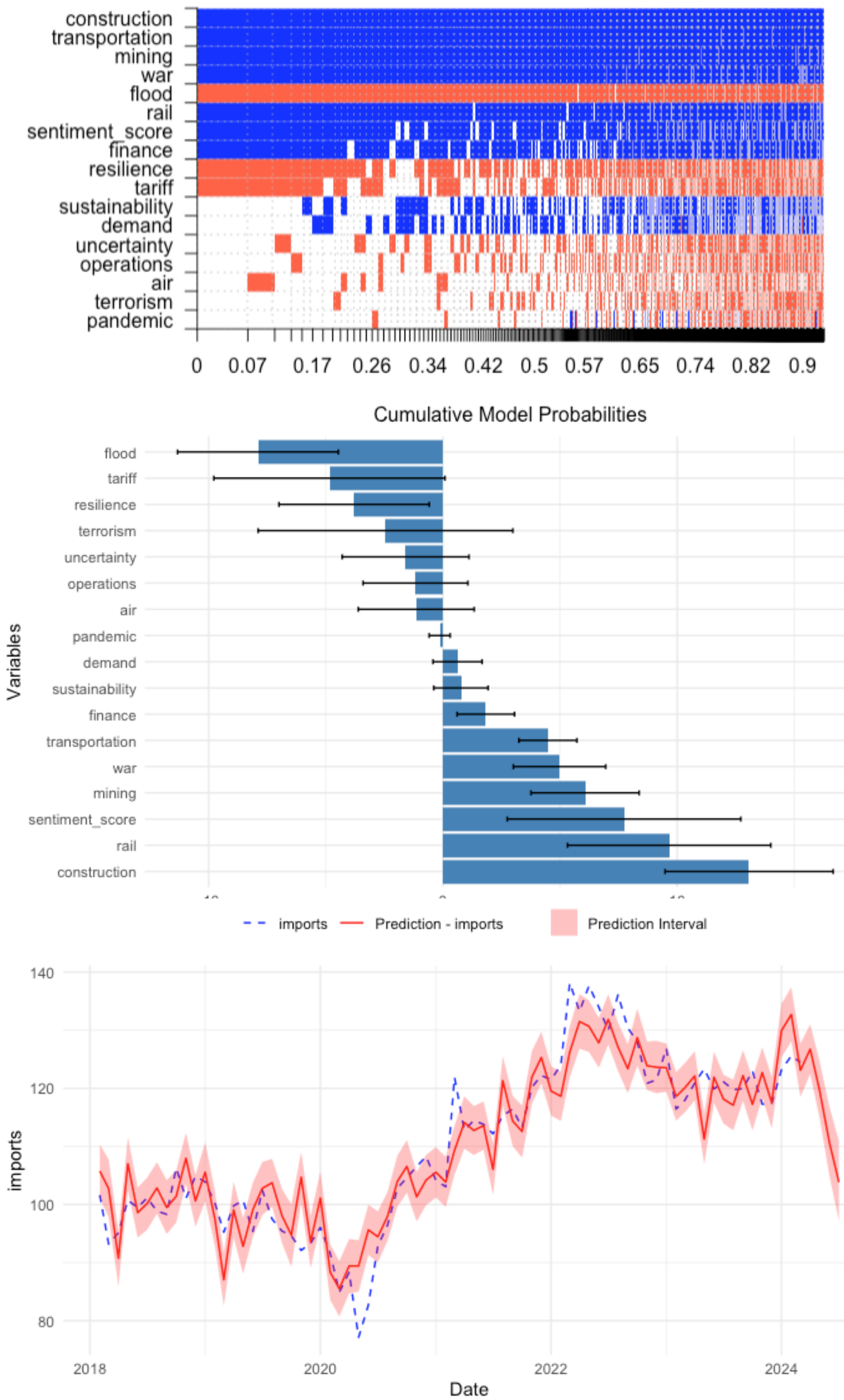


Equivalent R^2: 0.332

Figure 13. BMA model – Total non-farm employment – U.S. a) Model inclusion probabilities (495 models), b) marginal effects, c) outcome vs predicted values.

Sea Imports – U.S. (Seasonally Adjusted)

a) b) c)



Equivalent R^2: 0.872

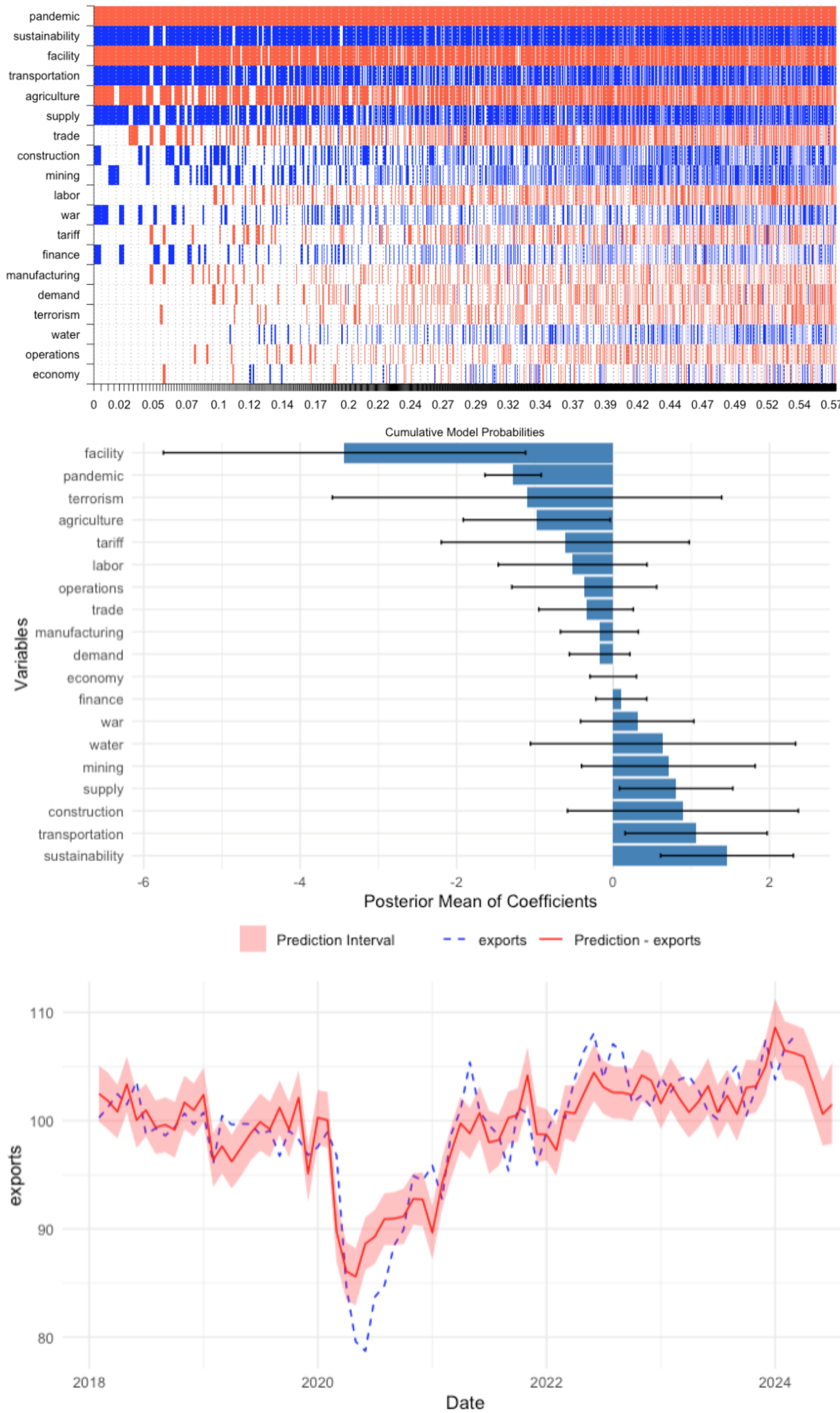
Figure 14. BMA model – Sea imports – U.S. a) Model inclusion probabilities (2000 models), b) marginal effects, c) outcome vs predicted values.

Sea Exports – U.S. (Seasonally Adjusted)

a)

b)

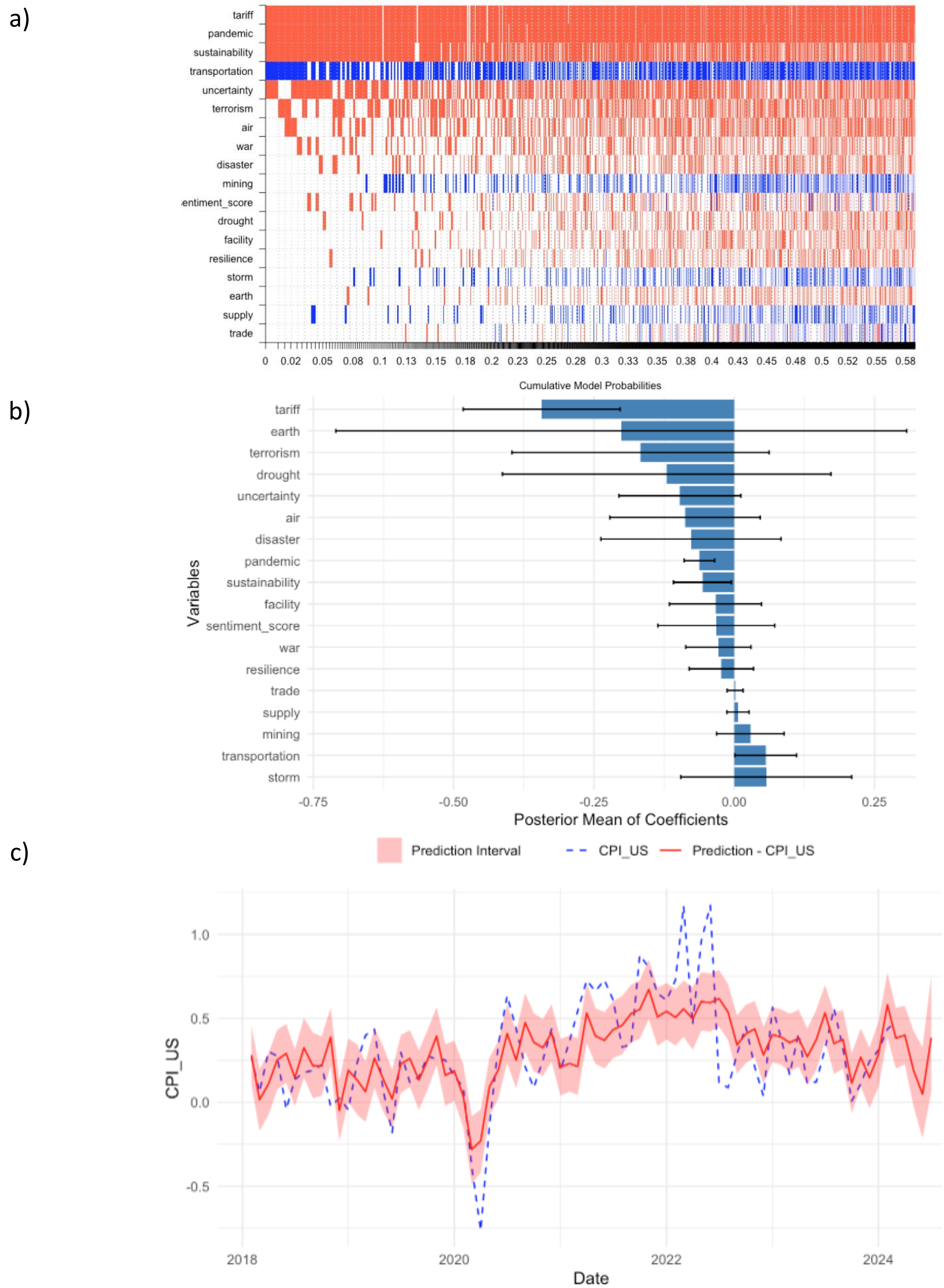
c)



Equivalent R^2: 0.758

Figure 15. BMA model – Sea exports – U.S. a) Model inclusion probabilities (2000 models), b) marginal effects, c) outcome vs predicted values.

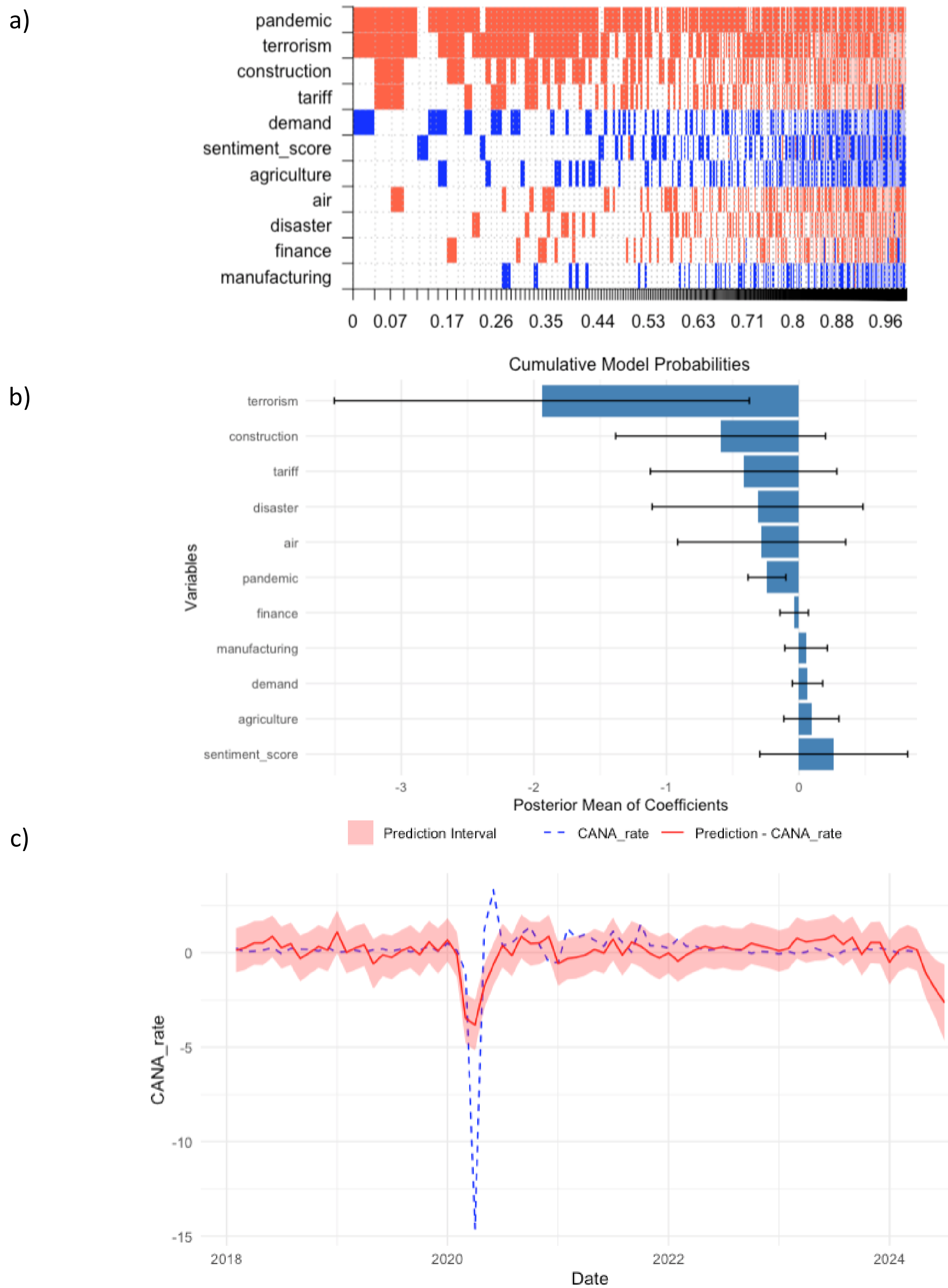
Consumer Price Index – U.S.



Equivalent R^2: 0.551

Figure 16. BMA model – CPI – US. a) Model inclusion probabilities (2000 models), b) marginal effects, c) outcome vs predicted values.

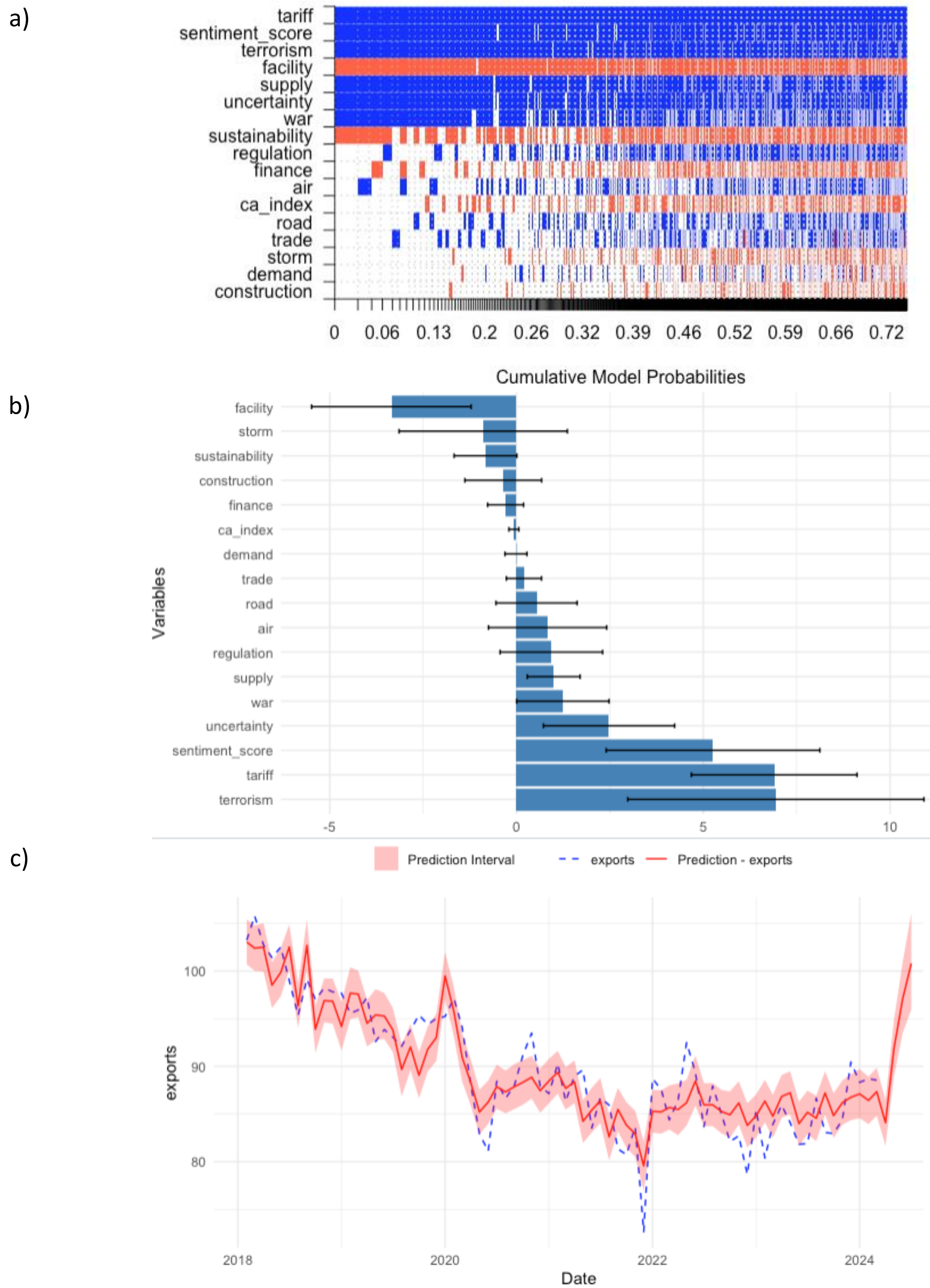
Total Non-farm Employment - California



Equivalent R^2: 0.371

Figure 17. BMA model – non-farm employment in California. a) Model inclusion probabilities (1314 models), b) marginal effects, c) outcome vs predicted values.

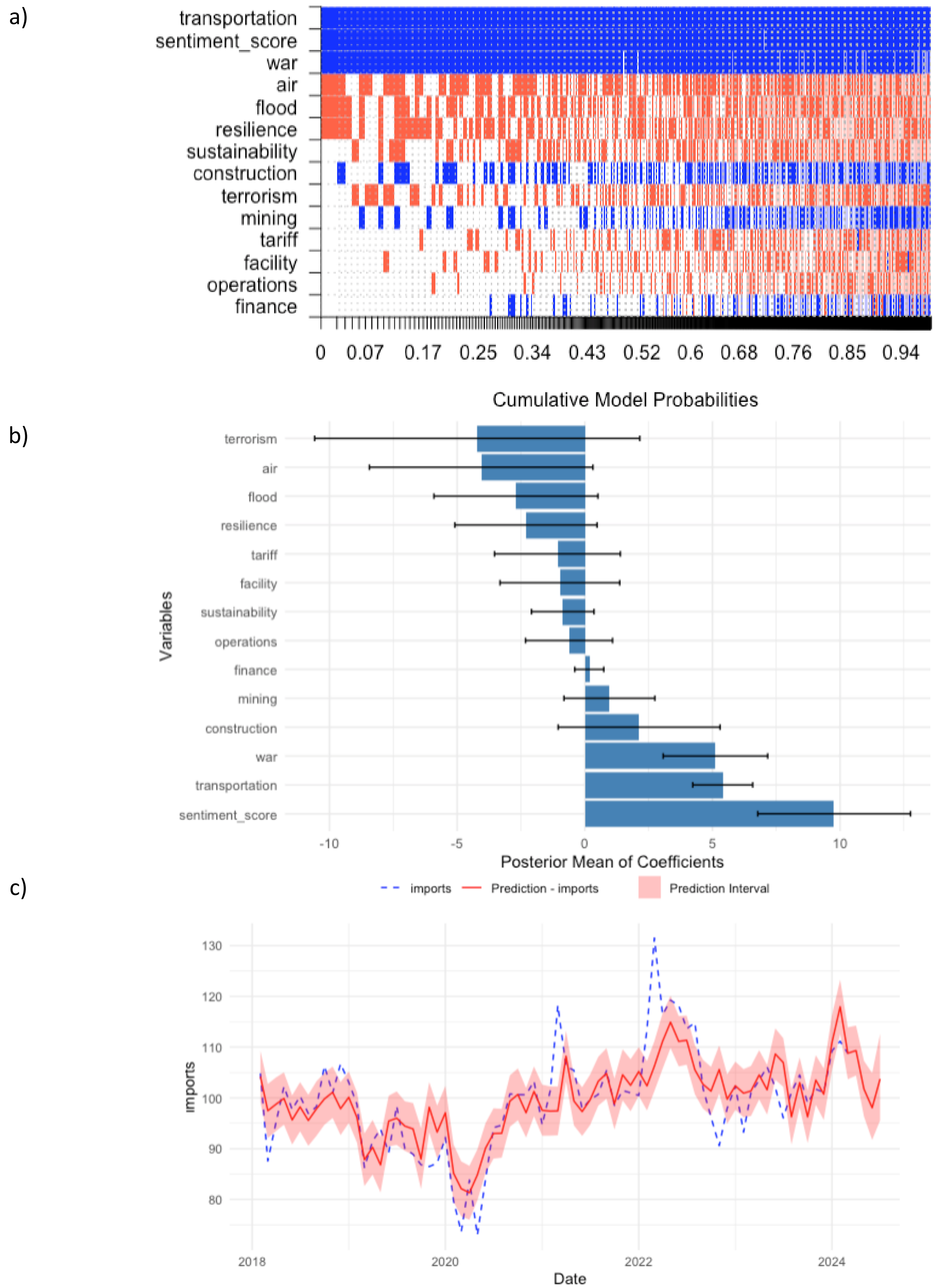
Sea Exports – California (Seasonally Adjusted)



Equivalent R^2: 0.832

Figure 18. BMA model – sea exports in California. a) Model inclusion probabilities (2000 models), b) marginal effects, c) outcome vs predicted values.

Sea Imports – California (Seasonally Adjusted)



Equivalent R^2: 0.652

Figure 19. BMA model – sea imports in California. a) Model inclusion probabilities (2000 models), b) marginal effects, c) outcome vs predicted values.

Appendix C – Empirical Validation in Task 2

The team empirically validated the performance metrics by analyzing the supply chain's recent disruptive events. The evaluation focuses on preliminary results for specific periods to assess the metrics' effectiveness and relevance, as well as those metrics' ability to help identify issues. These preliminary results can change as the team improves some techniques for enhanced tool efficiency. The initial evaluation helps gauge the tool's capability to identify and analyze supply chain disruptions across time, determine the coherence and usefulness of the results, and highlight areas where refinement and improvement are required to enhance the tool's performance and utility.

The evaluated periods include:

1. **COVID-19 (2020).** The team evaluated the performance of the metrics during the period marked by the COVID-19 pandemic. Analyzing news metrics and supply chain data during pandemic outbreaks allows for assessing the metrics' effectiveness in capturing supply chain risks. Figure 20 shows the resulting metrics (number of news articles for the specific topic and the sentiment analysis).

   The empirical results obtained during the COVID-19 event align with the observed trends and impacts on the supply chain. At the beginning of COVID-19 (March 2020), there was a notable surge in the frequency of news articles published daily, rising from an average of 10 to approximately 40 news articles per day. This increase in news coverage reflects heightened attention and interest in supply chain-related developments and disruptions amidst the pandemic. Also, in March 2020, the sentiment of news articles experienced a significant decrease. This decline in sentiment indicates a shift towards more negative tones in news coverage, likely reflecting concerns, uncertainties, and challenges supply chains faced in response to the pandemic's impacts. The word cloud reveals that some of the most relevant words during this period include pandemic-related terms such as "risk," "virus," "China," "test," and "cost." These keywords indicate the predominant themes and challenges affecting supply chains during the initial phases of the COVID-19 pandemic, including supply chain disruptions, health and safety concerns, global trade impacts, and cost implications.

   Finally, the geographic concentration of news articles underscores the significance of China and the United States as relevant locations during the COVID-19 pandemic. This concentration reflects the prominent roles of these countries during this event, especially in supply chain-related topics.

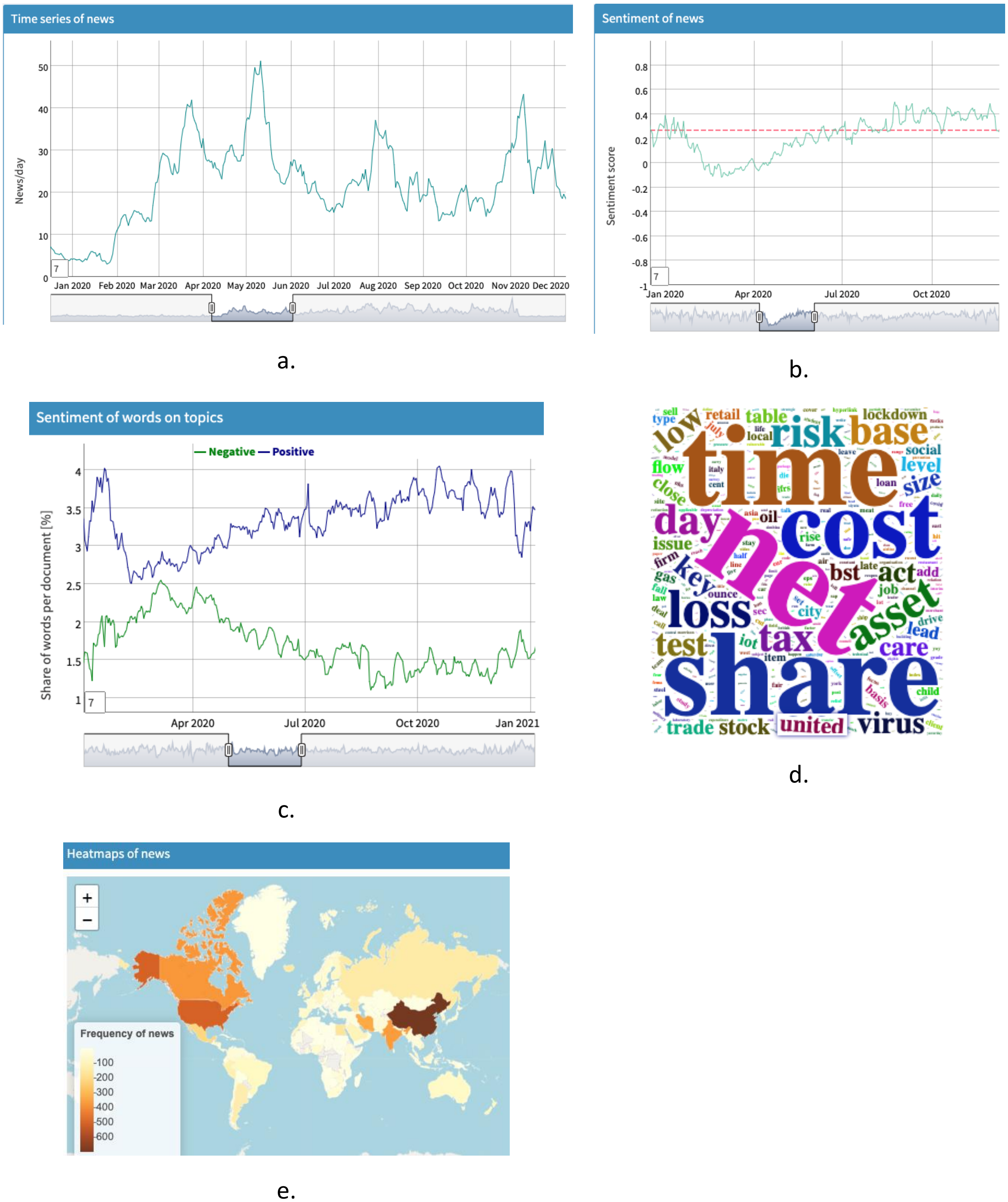


Figure 20. Results from Jan 2020 to Dec 2020

2. **Russian-Ukrainian conflict (Feb to May 2022):** Now, the metrics are evaluated during the occurrence of geopolitical events such as the Russian-Ukrainian war, the results of which are in Figure 21.

Pacific Southwest Region UTC
University Transportation Center

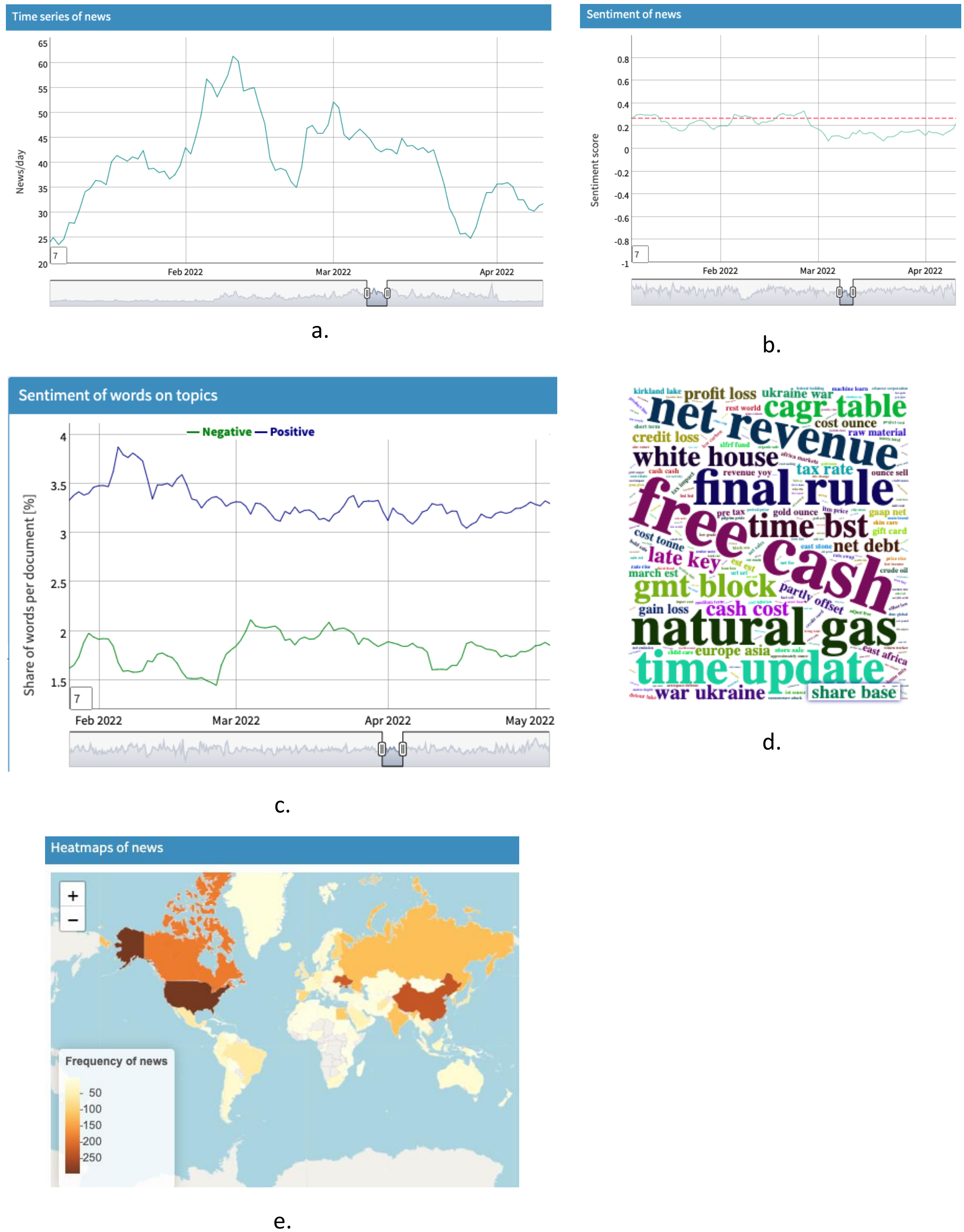
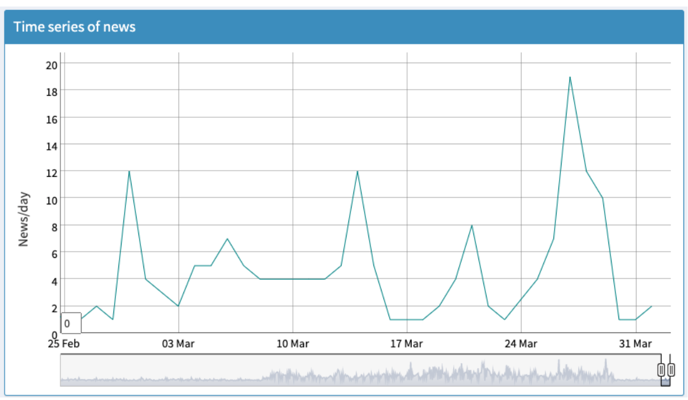


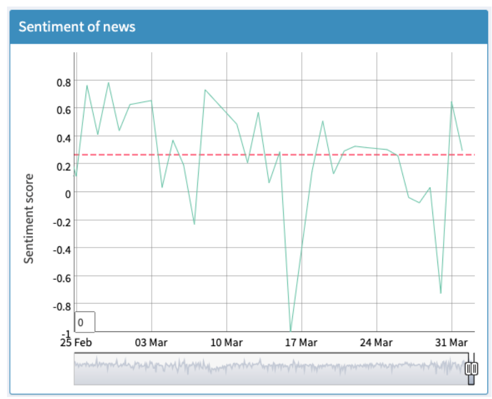
Figure 21. Results in Feb 2022 to Dec 2022

During this period, a similar trend was observed in the frequency of news coverage, indicating an escalation in the reporting of supply chain risks. Conversely, there was a corresponding decline in the sentiment of news articles, suggesting a shift towards more negative tones in the reporting during this timeframe. Additionally, the geographical concentration of news coverage shifted to focus on Ukraine, Russia, and the United States as key geographies where news articles were concentrated. The word cloud analysis revealed several key terms that were prominently featured during this period, including “natural gas,” “Ukraine war,” “White House,” and “profit loss.” These terms indicate the predominant themes and issues that dominated news coverage related to supply chain risks during this timeframe.

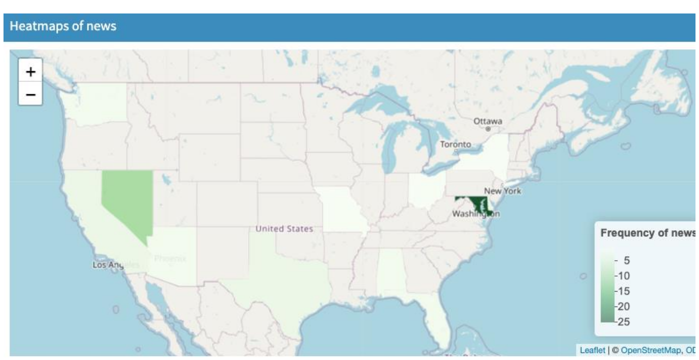
3. **Baltimore’s bridge collapse (March 2024):** The collapse of Baltimore’s bridge in March 2024 serves as a critical event for the empirical validation of supply chain metrics. This incident, being a key component for merchandise trade in the United States and the national supply chain, presents an opportunity to assess the effective
4. ness of the performance metrics in capturing and analyzing supply chain disruptions in quasi-real time. Figure 22 depicts the results.



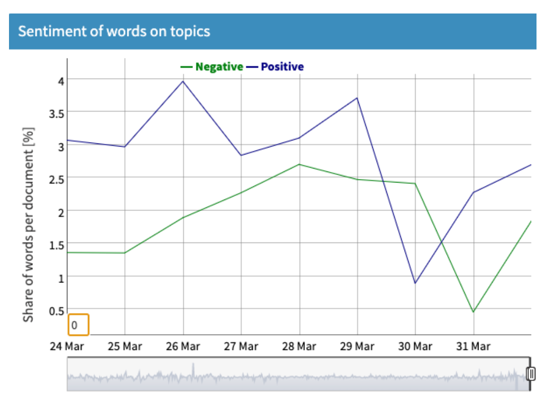
a.



b.



c.



d.

Figure 22. Results in March 2024

The empirical results following the collapse of Baltimore's bridge on March 26 reveal significant changes in key metrics related to supply chain risk, including a surge in the frequency of news articles following the bridge collapse. This increase in news coverage reflects heightened media attention and public interest in the event's impact on the supply chain. Additionally, the sentiment analysis of news articles indicates a predominant negative sentiment surrounding the supply chain following the bridge collapse. This is explained when reviewing the sentiment of words, which shows a significant increase in negative wording after March 26, surpassing the share of positive wording. This negative sentiment likely stems from concerns about transportation disruptions, logistical challenges, and potential economic consequences arising from the incident. Finally, the geographic concentration of news articles primarily focuses on the State of Maryland, where the bridge collapse occurred.