Trends of Economics: A Descriptive Text Analysis for Economic Literature

By Henry Liu *

Large scale complex text analysis for economics literature used to be infeasible until Natural Language Processing (NLP) techniques allow general cognitive ability for text. This study employs three NLP models to conduct a descriptive analysis of 74,292 abstracts from papers published in the top 30 economic journals between 1990 and 2023. The report document trends in methodology, writing style, and academic fairness in economic literature. To my knowledge, this report is the first to analyze economics literature in the sentence level.
JEL: A14, B23, C18
Keywords: Text Analysis, Economic Literature, NLP

Economic literature corpus, despite their large volume of more than 5 million journal articles, exist in a generally well-structured textual form. Thus, employing statistical techniques related to text analysis to distill key topics, trends, and paradigm shifts therein, is feasible once they are classified into components or categories.

These texts typically cover theories, empirical analyses, and policy discussions, with connections between articles sometimes obscure, requiring robust understanding and experienced practice to meaningfully engage in even classifying them. Considering the scarcity of economic researchers and potentially high opportunity costs of their time, categorizing them would be sometimes too laborious even for their research assistants.

Recent advancements in machine learning, notably in deep learning, have significantly enhanced and democratized text analysis capabilities. Marked by the release of Large Language Model (LLM) by OpenAI in the end of 2022, it has been making text analysis significantly cheaper, easier-to-use and more accessible. Tools, such as LLM and its derivatives, which attain their general text processing ability via mass scale next-token-prediction training, can undertake some of the necessary but laborious text-related tasks for economic researchers.

Researches in economic field quickly adapted it to their own research pipeline. One example is the introductory paper *Generative AI for Economic Research: Use Cases and Implications for Economists* (Korinek 2023), establishing a first public approval on using LLM for economic research in economics study.

* The author affiliates to Ludwig-Maximilians-Universität München and ifo Leibniz Institute for Economic Research. He appreciates the advice and helps from Prof. Dr. Mathias Bühler from Ludwig-Maximilians-Universität München, and Nikola Noske from Leibniz Institute for Economic Research. All flaws remain to the author.

I. Literature Review

A. Text Analysis Used in Economic Studies

Historically, NLP and machine learning have been used in economic research to create variables for regression models, acting as annotation tools (Ash and Hansen 2023). For instance, Elliott Ash, Germain Gauthier and Philine Widmer propose a unsupervised method to identify the narrative of politician for political economics (Ash, Gauthier and Widmer 2024).

Besides, at the ifo Institute, Anna Kerkhof employed Natural Language Processing techniques to assess the welfare impact on viewers following a reform on YouTube, aligning with the approach of generating variables from textual content (Kerkhof 2020). Sarah H. Bana analyzed job postings using word embedding techniques, explaining wage disparities with the derived vectors (Bana 2022). This method accounted for approximately 87% of the variation, illustrating the efficacy of converting textual information into numerical data for economic analyses.

A study by W. Walker Hanlon using NLP technology to examine the linkage between engineers during Britain's Industrial Revolution and local innovation patterns using data from *Oxford Dictionary of National Biographies*, which revealed the relation between the human capital of engineers and innovation pattern (Hanlon 2022).

Those methodologies exemplify the conversion of textual information into numerical data for analytical purposes. NLP facilitates the creation of variables from text, transforming it into vectors or scalars for quantitative analysis. This approach not only extends NLP's utility in economics beyond variable construction but also highlights its capability to extract insights from economic literature. The complementarity of traditional statistical methods and NLP, in terms of processing numerical and textual data respectively, underscores the significance of both in enhancing analytical efficiency and the broader application of text manipulation in economic research.

B. Text Analysis for Economic Literature

Beyond serving as instrumental tool during the pipeline of economic studies, statistical text analysis can also be used for analyzing economic literature themselves.

Several descriptive studies have provided insights into basic information of economic papers. Among them, "Nine Facts about Top Journals in Economics" by Card and DellaVigna offers a straightforward enumeration of yearly submissions, acceptances, article lengths, and other pertinent metrics within top economics journals (Card and DellaVigna 2013). They finds, for example, the average length of articles has tripled since the 1970s, and the average number of authors per paper has almost doubled from 1.3 to 2.3 over the top 5 journals.

The study conducted by Ana Rute Cardoso and others represents a pivotal contribution to the understanding of the international dynamics within economic research (Cardoso, Guimarães and Zimmermann 2010). Their work, while primarily applying statistical analysis rather than text-based approaches, identifies a trend that the shifting geographical distribution of publications within top economic journals became more equal. Specifically, they concluded a gradual decline in the proportion of top economic journals in North American, along with the increase in European journals. Their finding is enriched by this study, which draws a similar conclusion from the perspective of researchers' nationality.

Martin Paldam examined 3,415 economic publications from 1977 to 2017, categorizing them into theoretical, experimental, and empirical research (Paldam 2021). His analysis indicated a 26% decline in theoretical studies and a 15% increase in reduced-form analysis articles. This study employs a significantly larger dataset, enhancing the robustness of trend validation observed by Paldam. Unlike his approach, which utilized machine learning for classification, this research employs an expanded methodology to analyze trends in economic literature more comprehensively.

II. Data Acquisition and Preparation

A. Identification of Premier Journals

The initial step in compiling the dataset involved identifying the leading journals within the field of economics. This was achieved by consulting the rankings provided by the IDEAS organization. A list of the top 30 journals, according to the IDEAS ranking, was compiled. The rankings can be accessed at IDEAS official website[1].

B. Data Extraction from EconLit Database

Subsequently, the EconLit Database was utilized to extract metadata for the journals identified in the preceding step. EconLit, updated weekly and containing over 1.6 million records, covers economics literature from the past 130 years, sourced from 74 countries. Access to the database was obtained through Ludwig Maximilian University (LMU), a member institute. The database is accessible via LMU online library[2].

Using the advanced search feature of the EconLit Database, specifically the "source (SO)" field, metadata corresponding to the journals from the top 30 list was extracted. The results were exported in XML format, facilitated by an email delivery system providing a link for file download.

C. Conversion of XML Files to CSV Format

The acquired XML files were then converted into CSV format using an XML parser. This transformation was essential for standardizing the data structure, thereby facilitating further analytical procedures.

D. Consolidation of Data

The individually converted CSV files were amalgamated into a singular dataset. This dataset, encompassing columns such as journal title, ISSN, publication date, volume, issue, article title, authors, affiliations, subjects, abstract, publication type, language, and URL, serves as the basis for subsequent analyses. This comprehensive dataset encapsulates metadata of articles published in the leading economics journals, ready for in-depth exploratory analysis.

The detailed steps outlined above, from the identification of journals to the consolidation of data, ensure a clear and replicable process for dataset compilation, adhering to formal academic documentation standards. This concludes the data preparation phase.

[1] IDEAS website is https://ideas.repec.org/top/top.journals.all.html
[2] LMU online library is https://web-p-ebscohost-com.emedien.ub.uni-muenchen.de/ehost

III. Methodology

In this report, we train and deploy three NLP models for the analysis. The first model, Word2Vec, is utilized for computing word embeddings, enabling the identification of semantically similar words for robust feature engineering. The second model is a sentence classification model, which facilitates the categorization of sentences within abstracts, thereby refining our analysis. The third model predicts an individual's nationality based on their name, serving as an instrumental tool for examining the composition of authors in economic literature across different periods.

A. Word2Vec Model for Word Embedding

The word2vec methodology, developed by Tomáš Mikolov and his team at Google in 2013, stands as a pivotal advancement in the field of NLP (Mikolov et al. 2013). It involves deriving vector representations for words, encapsulating both their semantic meaning and contextual usage. By analyzing text within a substantial corpus, word2vec models are trained to approximate these vector representations. Upon training, these models gain the ability to identify synonyms and propose words to complete sentences, showcasing the depth of understanding word2vec algorithms achieve in capturing linguistic nuances.

The Word2Vec model, utilized for computing word embeddings, identifies semantically related words to improve linguistic analysis. By selecting a keyword as an anchor, it finds related keywords to assess if an article pertains to specific themes or methodologies. This approach, has been adapted in sociology research in order to effectively models societal phenomena. "The Geometry of Culture" by Kozlowski, Taddy, and Evans exemplifies the application of Word Embedding in analyzing the evolution of societal stereotypes, such gender bias and political tendency. Word2Vec, by transforming words into vector representations, captures semantic relationships, leveraging the context to predict word meanings.

The word vectors generated by Word2Vec capture rich semantic relationships, including synonyms, antonyms, and words interchangeable in certain contexts. The distance between vectors, such as cosine similarity, measures semantic similarity, positioning similar or related words proximately in the vector space.

By converting words into vectors, Word2Vec offers a powerful tool for Natural Language Processing (NLP) tasks, allowing machines to comprehend the semantic content of vocabulary. This capability facilitates applications in text classification, sentiment analysis, and machine translation, among others. In our context, the objective is to identify key words and their semantic neighbors, constructing a thematic 'bag of words' for analysis.

B. Sentence Classification Model

The abstracts of economic papers, particularly those published in high-standard journals, often adhere to specific formatting requirements. The patterns entailed by these requirements facilitate the classification of sentences within the abstracts, making the task more straightforward. ChatGPT-4 serves as the initial annotator for sentence classification, utilizing its text analysis capabilities (Achiam et al. 2023). The annotations generated by ChatGPT-4 are then used for fine-tuning a pre-trained BERT model, which allows it to transfer its pre-trained knowledge into classifying the sentences in abstract.

Due to budgetary constraints, I do not choose to entirely rely on use ChatGPT to classify all sentences, which is entirely possible otherwise. The appendices detail the code and prompts for result reproducibility.

Utilizing ChatGPT-4 for Data Annotation

For our fine-tune sample, we extracted 300 abstracts from papers published by the American Economic Reviews and ifo Institute, employing ChatGPT-4 as the annotator. The model, identified as GPT-4-1106-preview, was tasked with classifying each sentence of an abstract into one of four categories: background sentence, method sentence, result sentence, and other sentence. The strategy starts with segmenting the abstracts into individual sentences by SpaCy, a Python module specialized at natural language processing, each prefixed with a numerical identifier, and classifying them according to the specified categories (Honnibal et al. 2020).

The system prompt utilized for this task is as follows:

> You are a sentence classifier. The input is a list of sentences from the abstract of an academic paper. Your task is to determine which category a sentence belongs to:
>
> "Background": What others have done, what is the problem;
>
> "Method": What we do, what we use, how we analyze;
>
> "Result": What we find or show, what is the implication, a statement presenting the result;
>
> "Other": None of the above.
>
> Output in JSON format: {1: "xxx", 2: "xxx", 3: "xxx", ...}

It is important to note that the chosen prompt operates in a zero-shot manner, based on ChatGPT's training on vast amounts of text to establish clear definitions and representations for each category (Gilardi, Alizadeh and Kubli 2023).

The annotation process was completed within minutes, suggesting that expanding the training set size is feasible by merely increasing the sample quantity. A preliminary manual review of the classification results did not reveal any inappropriate categorizations. The training dataset will be included in the supplementary materials attached to the paper.

The BERT Model Architecture

The study utilizes the BERT model, an open-source framework developed by Google in 2018, renowned for its bidirectional Transformer architecture. BERT embodies an encoder-only architecture within the transformer model framework (Devlin et al. 2018). Its architecture is delineated into three principal components: a embedding module, a list of transformer encoders and a un-embedding module.

The embedding module, which translates an array of one-hot encoded tokens into a corresponding array of vector representations. This process facilitates the transition from discrete token representations to continuous vector spaces, capturing the semantic and syntactic nuances of the tokens.

A series of transformer encoders, constituting the core of the BERT architecture. These encoders iteratively refine the vectors through a sophisticated mechanism of self-attention and feedforward networks, enhancing the representation of each token by integrating information from its context within the text (Vaswani et al. 2017).

The un-embedding module, designed to revert the final array of representation vectors back into one-hot encoded tokens. While crucial for the pretraining phase of BERT, wherein the model learns to predict masked tokens and next sentences, this module is frequently omitted in downstream applications. In such scenarios, the representation vectors produced by the encoder stack are utilized directly. These vectors serve as input to specialized models tailored to specific NLP tasks, leveraging the rich contextual embeddings generated by BERT for tasks like sentiment analysis, question answering, and more.

Fine-Tuning The BERT Model

For specific natural language processing applications, such as text classification in our case, BERT's pre-training allows for subsequent fine-tuning to tailor its capabilities to particular tasks. This research focuses on adapting BERT for text classification, employing ChatGPT-4 generated data for fine-tuning (Wei et al. 2021).

The fine-tuning procedure was executed on a MacBook Air with an M2 chip, taking approximately 10 hours. For larger datasets, cloud computing resources could offer more efficient fine-tuning processes.

Inference Time Modification

Post-fine-tuning, a challenge encountered during sentence classification of a vast dataset was the excessive length of some sentences. Given the $O(N\hat{2})$ complexity of Transformers, like BERT, inference times far exceeded expectations. To mitigate this, a keyword extraction method was employed, using the SpaCy package for word tagging to extract subjects, verbs, and direct objects (Honnibal et al. 2020). This approach effectively shortened sentence lengths while retaining the core information, thus resolving the issue without significantly impacting classification accuracy. In fact, accuracy improved for some longer sentences, likely due to the smaller size of this particular BERT model variant's limited capability in handling longer sequences.

C. Name-to-Nationality Prediction Model

Model Application and Dataset Description

In developing the Name-to-Nationality prediction model, we employed a similar methodology and architecture, specifically leveraging a BERT model fine-tuned for predicting nationality based on names. This predictor has been encapsulated into a package, facilitating ease of use for any interested parties. It is pertinent to highlight the source of the data for the Name-to-Nationality model. The dataset was compiled by a GitHub user who extracted the entirety of Wikipedia and identified nationality information from the bottom of each Wikipedia page, successfully creating a comprehensive dataset (Park 2020).

The advantage of such an extensive dataset is its inclusivity of virtually all common human names. Consequently, even if overfitting occurs, it is unlikely to pose practical issues in the majority of applications, given the dataset's breadth. During training, an 80-20 split was applied for training versus testing and validation. The model version with the lowest validation loss was selected for packaging and inclusion in the subsequent name classification pipeline.

Discussion of Biases and Limitations

However, the dataset does have a significant limitation: the representation of nationalities on Wikipedia is not proportionate to the global population distribution, with a predominant focus on individuals from the United States. This skew could introduce biases in the model, particularly under-representing countries and individuals with lower internet presence. While any specific tests can be conducted using the packaged model, these biases are not necessarily detrimental within the context of this report for two reasons. First, many prominent economists and their works are likely included in the training set, ensuring high accuracy for predictions concerning these individuals. Even if a name gives an impression of a nationality that does not match the actual fact, the model might still classify it correctly. Second, given the dominant position of the United States in the field of economics—both in terms of research institutions and publication venues—the dataset's bias towards American data might inadvertently serve as an advantage in prediction tasks. However, this observation is highly specific and localized; caution is advised when applying the model in other contexts.

IV. Findings: Trends in Economic Research

Through a comprehensive analysis of economic literature from various perspectives, a wealth of information has been collated, leading to significant insights into the field. This section elucidates changes across three dimensions within economic research: trends in methodologies, shifts in writing styles, and variations in the composition of authors over time.

A. Changes in Methodologies

The findings indicate a discernible shift in economic research methodologies over the past two decades, with a pronounced inclination towards empirical research and causal inference. This transition reflects a broader evolution in the discipline, emphasizing the importance of establishing causality and leveraging empirical data to inform economic theories and policies. The adoption of more sophisticated econometric models and the increased availability of data have facilitated this shift, enabling researchers to address complex economic questions with greater precision.

The analysis of methodological trends within economic literature was enhanced by employing the previously discussed sentence classifier to categorize sentences in abstracts into four distinct groups. From this classification, sentences related to methodology were extracted for further analysis. A bag-of-words (BOW) counting method was utilized to identify different topics of interest. This indicator is defined as one when any one of the 5 words exists in the text, and zero otherwise, mathematically:

$$I = \begin{cases} 1 & \text{if } \exists\, w \in \{w_1, w_2, w_3, w_4, w_5\} \text{ such that } w \in \text{Text} \\ 0 & \text{otherwise} \end{cases}$$

Specifically, five key topics were targeted: model, theory, effect, data, and identification. By employing a pre-trained Word2Vec model on these abstracts, the most closely related four keywords for each topic were identified. An indicator was defined such that if any of the five words appeared in a target sentence, it was deemed to contain that topic.

The trends observed in A1 and A2 can be bifurcated into two groups: stable and ascending. Model and theory belong to the stable group, exhibiting relatively unchanged prevalence over the past two to three decades. In contrast, effect, data, and identification have seen substantial increases in their representation over the same period.

The occurrence rates of effect and data escalated from approximately 10% in 1990 to nearly 25% by 2023, showcasing similar trajectories. This similarity underscores a shift towards more empirical research in economics and a greater emphasis on the pivotal role of data in addressing economic inquiries.

As shown in A2, identification's representation grew from about 1.5% in 1990 to close to 7% in 2023, reflecting a significant increase and mirroring the intuitive perceptions of empirical economists. Notably, this trend only gained widespread momentum after the year 2000, potentially influenced by the widespread adoption of personal computers and the launch of Windows XP, which facilitated diverse regression analyses and more sophisticated data processing. However, this conjuncture lacks causal identification.

Regarding the stable topics in A1, theory has maintained a consistent state over the past 30 years, fluctuating between 20% to 25%. In comparison, the topic of model has seen a notable increase, rising from just over 20% in the 1990s to about 28% today. Despite this growth, it does not match the rate of increase observed in the other topics on a relative scale. Interestingly, even though the growth rate of model and theory is not as pronounced as the other three topics, they remain dominant in terms of absolute values. For instance, model consistently outperforms the other groups in average mentions across all years, and theory surpasses the other three in most years. However, post-2020, the proportions of theory align closely with effect and data, ranging between 20% to 25%, indicating a competitive equilibrium among these topics.

B. Trends in Writing Styles

Over recent decades, not only have the methodologies employed by economists undergone substantial changes, but the characteristics of their writing have also evolved significantly. Among the most notable shifts is the tendency for abstracts to become increasingly lengthy, a trend that reflects broader changes in the structure and detail of economic research papers.

Longer Abstracts

Abstracts have shown a marked increase in word count in A3, escalating from approximately 100 words in the early 1990s to about 150 words today, representing a 50% increase. However, this growth in abstract length pales in comparison to the expansion observed in the main body of papers, which has increased several-fold in recent years (Card and DellaVigna 2013). Several factors drive this trend towards longer abstracts, notably the increased incorporation of numerical data. This suggests that authors are more frequently detailing empirical results, such as the specific percentage impacts of policies, within their abstracts.

Word-level Features

The analysis extends to word-level features such as word count, the number of numerals, counts of capital letters, and punctuation marks. A significant finding in A5 is the increase in the use of numbers within abstracts, indicative of a shift towards presenting more precise empirical results.

Moreover, A6 reveals a substantial increase in the number of capital letters used per sentence, rising from an average of 1.75 in the early '90s to 2.3 in recent years. Discounting the capital letter typically used at the beginning of each sentence, the early '90s had approximately 0.75 capitalized nouns per sentence, which has grown to about 1.3 in the present day. This near 100% increase is a testament to the growing specialization and formalization within the field, with research increasingly built upon empirical foundations, employing specialized methodologies, or mentioning specific methods, concepts, countries, or individuals.

C. Sentence-level Features

On the sentence level, as indicated by A9, the prevalence of longer sentences has increased, from an average of 1.2 long sentences (defined as those exceeding 30 words) per abstract in the early 90s to nearly 1.8 today. This 50% increase in longer sentences often associates with more skilled writers. This trend may reflect the elevated standards of writing in top-tier economic journals, indicating an overall improvement in the quality of economic research publications.

D. Trends in Author Composition

The composition of authors contributing to economic literature has become increasingly diverse. Illustrated by A15, this diversification reflects the global expansion of the economic research community and the growing inclusion of underrepresented groups. The democratization of access to research platforms and collaborative networks, alongside efforts to promote diversity within academia, has contributed to this trend. The broadening of authorship not only enriches the field with a variety of perspectives but also enhances the applicability and relevance of economic research across different contexts and populations.

In summary, these findings underscore significant transformations in the landscape of economic research, marked by methodological advancements, evolving writing conventions, and increasing diversity among researchers. These trends collectively contribute to the dynamism and progress of the discipline, ensuring its continued relevance and impact in addressing contemporary economic challenges.

V. Discussion

A. Higher Focus on Causal Inference

Among the three main findings of this report, one notable trend is the growing focus on empirical research and the quest for causal relationships. This evolution suggests that forthcoming research will likely use large datasets and sophisticated statistical methods to establish and validate causal connections among economic variables. Consequently, researchers might need to enhance their proficiency in statistical and data analysis techniques to remain at the forefront of this shift.

B. Formal Writing and Specialization

Writing becomes a basic requirement. The evolution in economic literature also reflects changes in writing styles, marked by lengthier abstracts, more comprehensive data presentation, and an increased reliance on specialized terminology. These trends indicate a move towards more complex and nuanced economic discourse, necessitating that future economists not only master research methodologies but also excel in conveying intricate findings in an accessible and precise manner.

C. Greater Diversity

Towards greater author diversity is perhaps one of the most important progress in economics field. It moves from a predominantly US-centric authorship to a more globally representative group. Incorporating diverse perspectives in economic research, which could enrich understanding of global economic dynamics and relations. The increasing participation of authors from varied cultural and economic backgrounds may lead to a broader focus on international economic issues, fostering a more inclusive and comprehensive body of economic knowledge.

D. A Scalable Framework for Economic Content Analysis

Textual research on economic literature remains largely fragmented. Economic literature analysis often lacks a cohesive overarching theory, with each researcher attempting to analyze from different angles. This report is no exception. However, it explores this direction by attempting to construct a framework for analyzing economic papers, distinguished by Background, Method, and Result sections. While not innovative in itself, the contribution of this report lies in providing a scalable framework that makes large-scale literature organization feasible.

VI. Conclusion and Future Directions

A. Summary

This report aims to provide a framework of descriptive text analysis for economic literature in sentence level. I propose and implement three NLP-based models to conduct this descriptive analysis at sentence levels based on more than 700,000 journal article abstracts from top 30 economic journals.

Three trends in economic literature from different perspectives are documented: authorship has now much nationality-balanced representation than 1990; the writing style has been moving to a direction longer and more connected to terms and works from other researchers; finally, the methodology economics papers applied has been tiled to more data and causality-focused.

B. Economic Sentence Classification Dataset

As a byproduct, this work has also made available the first sentence-level economic literature classification dataset, which can serve various purposes. For instance, researchers might leverage text embedding-based search tools to rapidly review decades of economic literature on specific topics, facilitating a more efficient exploration of existing findings.

C. Future Directions

Besides a descriptive study, text analysis toolkit can also become a practical instrument facilitates scientific inquiry. For example, based on the sentence classification dataset attached to this report, researchers could use text embedding-based search tools to navigate through economic findings on specific domains; this function will be shortly available on the GitHub repository for this project[3]. This approach could be integrated with language models to create interactive economic research-focused chatbots, serving as a new infrastructure for future economists.

[3]The GitHub Repository for this report is https://github.com/Henry-Louis/TextAnalysisEconomicLiterature

REFERENCES

Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. “Gpt-4 technical report.” *arXiv preprint arXiv:2303.08774.*

Ash, Elliott, and Stephen Hansen. 2023. “Text algorithms in economics.” *Annual Review of Economics*, 15: 659–688.

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2024. “Relatio: Text semantics capture political and economic narratives.” *Political Analysis*, 32(1): 115–132.

Bana, Sarah H. 2022. “Work2vec: using language models to understand wage premia.” *Unpublished Manuscript.*

Card, David, and Stefano DellaVigna. 2013. “Nine facts about top journals in economics.” *Journal of Economic literature*, 51(1): 144–161.

Cardoso, Ana Rute, Paulo Guimarães, and Klaus F Zimmermann. 2010. “Trends in economic research: An international perspective.” *Kyklos*, 63(4): 479–494.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “Bert: Pre-training of deep bidirectional transformers for language understanding.” *arXiv preprint arXiv:1810.04805.*

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. “ChatGPT outperforms crowd workers for text-annotation tasks.” *Proceedings of the National Academy of Sciences*, 120(30): e2305016120.

Hanlon, W Walker. 2022. “Bridging Science and Technology During the Industrial Revolution.” Working Paper.

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. “spaCy: Industrial-strength Natural Language Processing in Python.”

Kerkhof, Anna. 2020. “Advertising and content differentiation: Evidence from YouTube.”

Korinek, Anton. 2023. “Generative AI for economic research: Use cases and implications for economists.” *Journal of Economic Literature*, 61(4): 1281–1317.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient estimation of word representations in vector space.” *arXiv preprint arXiv:1301.3781.*

Paldam, Martin. 2021. “Methods used in economic research: An empirical study of trends and levels.” *Economics*, 15(1): 28–42.

Park, Kyubyong. 2020. “name2nat: a Python package for nationality prediction from a name.” *https://github.com/Kyubyong/name2nat.*

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion

Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." *Advances in neural information processing systems*, 30.

Wei, Jason, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. "Fine-tuned language models are zero-shot learners." *arXiv preprint arXiv:2109.01652.*

Appendix

Figure A1. Share of Papers Containing Different Methodology Over Years

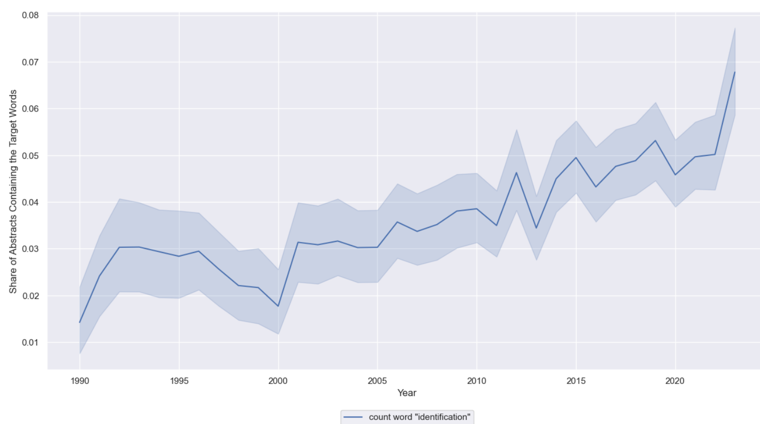

Figure A2. Share of Papers Containing Identification Over Years

Figure A3. Trend of Word Count

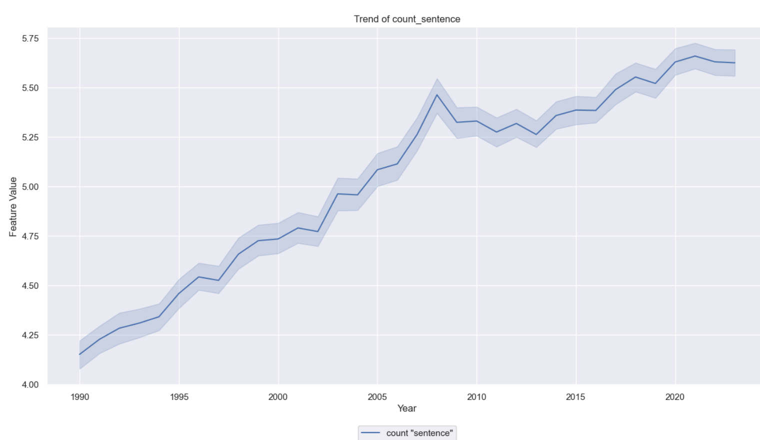

Figure A4. Trend of Sentence Count

FIGURE A5. TREND OF NUMBER COUNT

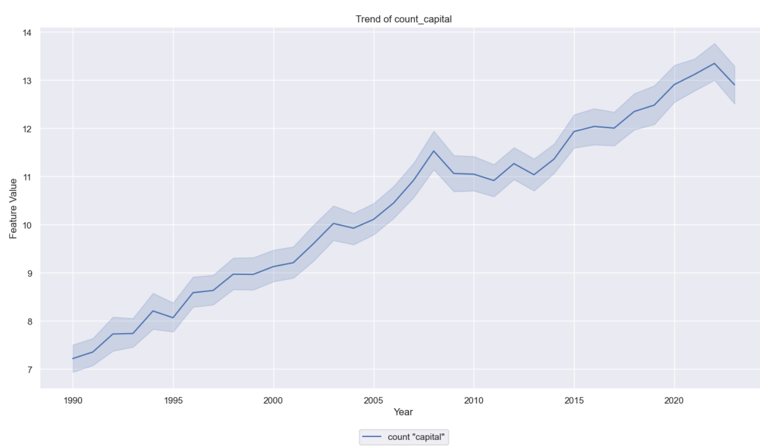

FIGURE A6. TREND OF CAPITAL LETTERS

Figure A7. Trend of Punctuation

Figure A8. Trend of Short Sentence Count

Figure A9. Trend of Long Sentence Count

Figure A10. Trend of Sentence Length Standard Deviation

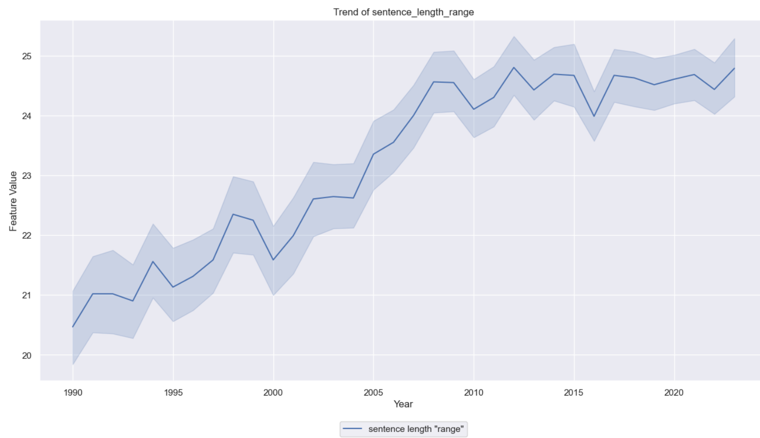

Figure A11. Trend of Sentence Length Range

Figure A12. Trend of Word per Sentence

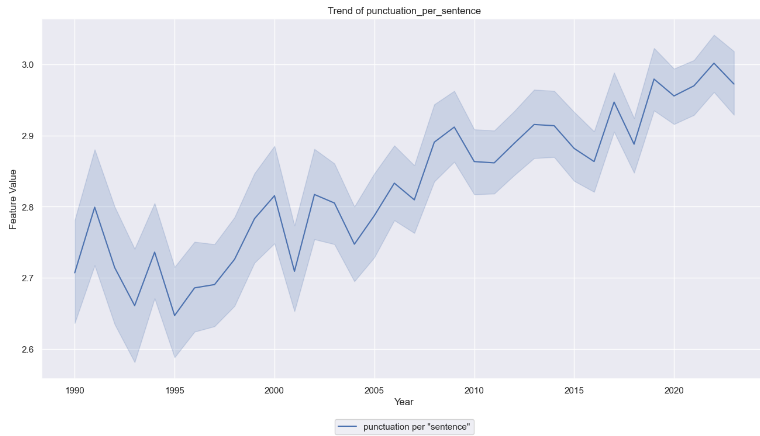

Figure A13. Trend of Punctuation per Sentence

Figure A14. Trend of Capital Letters per Sentence

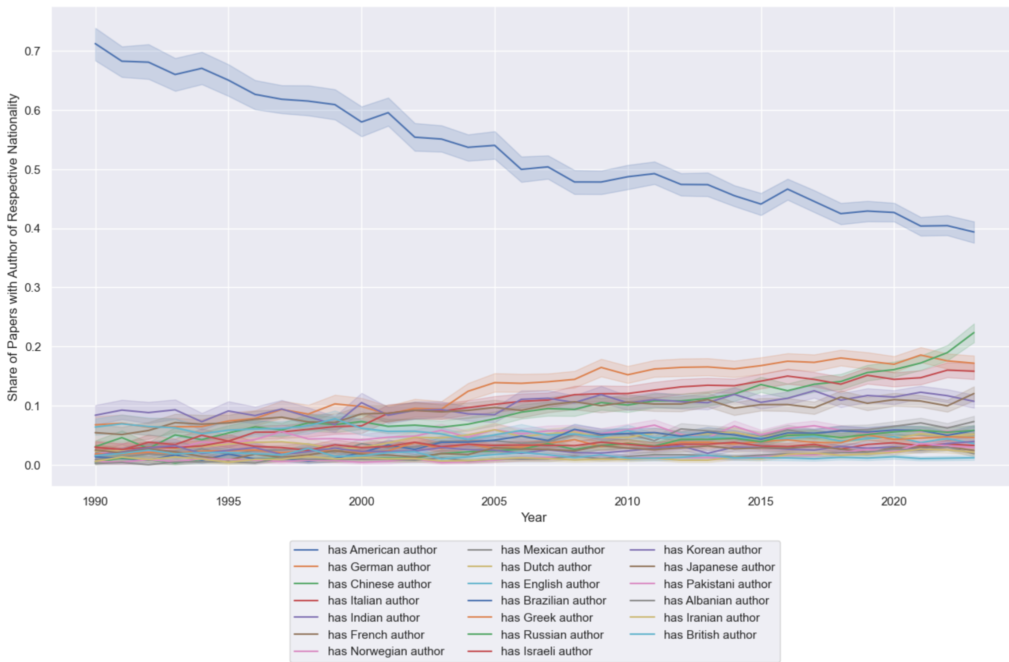

Figure A15. Share of Paper Having Author of Respective Nationality

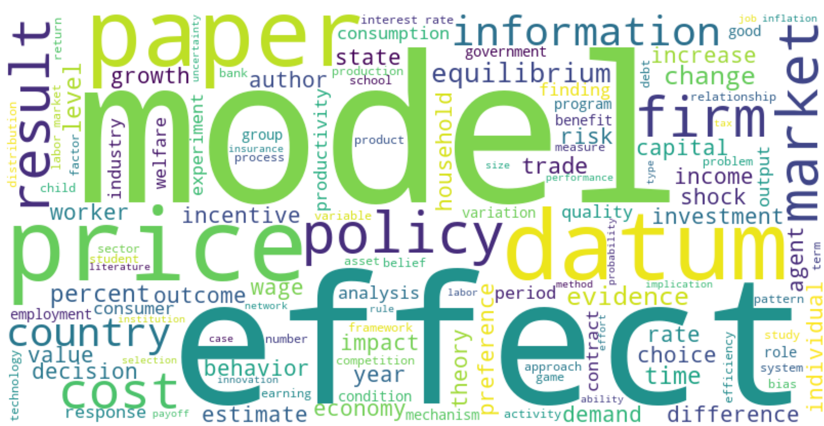

Figure A16. Noun Word Cloud for American Economic Reviews Abstract

Figure A17. Verb Word Cloud for American Economic Reviews Abstract f