Attraction is All You Need: The Impact of Eye-Catching Titles on Publication Success and Citation Rates in Economics

SHEN Zekai*

First Version: October 2024
This Version: October 2024

Abstract

We investigate the causal impact of title attractiveness on publication outcomes and citation counts in economics. Using a novel two-stage methodology combining human evaluation with machine learning, we analyze 347,197 papers published in 328 economics journals from 2000 to 2022. Papers with attractive titles are published in journals with ratings 0.120 points higher and receive 1.925 more citations on average. The effect is most pronounced for mid-tier journals and persists after controlling for author, institutional, and journal characteristics. We find that different large language models exhibit varying capabilities in predicting publication success versus citation impact, suggesting distinct mechanisms through which title attractiveness affects academic outcomes. Our findings contribute to the literature by providing causal evidence on how title characteristics influence academic success and demonstrating the potential of machine learning in analyzing subjective paper features at scale.

*NUS Business School, National University of Singapore; zekai_99@nus.edu.sg

I. Introduction

In the realm of academic economics, where the adage "publish or perish" holds increasing sway, the path to publication has become increasingly challenging. Over the past two decades, top economics journals have seen a substantial rise in submissions, with acceptance rates plummeting from around 15% in the early 2000s to below 7% in recent years (Card and DellaVigna, 2013). Concurrently, the average time from submission to publication has nearly doubled, now often exceeding two years (Hadavand, Hamermesh, and Wilson, 2024). In this fiercely competitive landscape, a well-received, highly cited paper can significantly enhance an economist's tenure prospects(Heckman and Moktan, 2020) and scholarly reputation(Hamermesh, 2018).

As the standards for exceptional contributions in economics continue to rise, researchers face mounting pressure to optimize every aspect of their scholarly output. While research quality remains paramount, a growing body of literature suggests that various ancillary factors, particularly those related to presentation, significantly influence publication success and citation impact. These presentation-related factors span a wide range of elements. Article length has been shown to affect both publication chances and citation rates(Hasan and Breunig, 2021). The clarity and style of writing can impact a paper's reception and subsequent influence (Feld, Lines, and Ross, 2024). Even the choice of coauthors can play a role in a paper's success, with factors such as coauthor networks (Ductor and Visser, 2022) and gender composition (Bransch and Kvasnicka, 2022) showing significant effects. Strategic decisions also matter, including the timing of submission(Ma, Li, Guo, and Si, 2019) and post-publication dissemination strategies, such as the use of social media for promotion (Chan, Önder, Schweitzer, and Torgler, 2023). Intriguingly, even seemingly peripheral elements like abstract readability have been shown to affect a paper's reception (McCannon, 2019). These studies collectively highlight the multifaceted nature of factors influencing academic success beyond the core quality of research itself.

Among these factors, the paper's title—the most concise yet visible component of academic work—warrants particular attention as the initial point of contact between research and its potential audience. Despite representing a minimal time investment in the overall research process, emerging evidence suggests that a paper's title may exert a disproportionate influence on both publication outcomes and subsequent citations in economics (Guo, Ma, Shi, and Zong, 2018; Bramoullé and Ductor, 2018; Gnewuch and Wohlrabe, 2017).

Existing research in economics has primarily focused on the correlation between observable title characteristics—such as length and the presence of nonalphabetical characters—and publication and citation outcomes. These studies generally find that shorter titles (Bramoullé and Ductor, 2018) and those incorporating non-alphabetical characters(Gnewuch and Wohlrabe, 2017) are associated with improved publication success and higher citation rates. However, this body of work has two primary limitations. First, it fails to provide a unified explanation for the seemingly disparate effects of various title attributes. Second, and more critically, it does not establish a causal link between title characteristics and academic outcomes, leaving open the possibility that these correlations are driven by unobserved factors.

We posit that the "attractiveness" of a title serves as the primary channel through which title characteristics influence a paper's academic trajectory. This hypothesis offers a unifying framework to understand the impact of titles on publication and citation outcomes, addressing a significant gap in the current literature. Our approach is grounded in behavioral economics theories, particularly those related to decision-making under limited attention(Sims, 2003) and framing effects(Tversky and Kahneman, 1981).

We propose that title attractiveness plays a crucial role in both the publication and citation processes, operating through distinct but interrelated mechanisms. In the publication process, attractive titles may capture the limited attention of editors and reviewers, increasing the likelihood of further consideration. They may also create a positive framing effect, potentially influencing the overall evaluation. In the citation process, attractive titles may enhance memorability, making papers more likely to be recalled and cited, consistent with the availability heuristic (Tversky and Kahneman, 1973). Furthermore, attractive titles may increase reader engagement, leading to more thorough processing of the paper's content, in line with the elaboration likelihood model (Petty and Briñol, 2011).

To test these hypotheses, we construct a comprehensive dataset of 347,197 papers published between 2000 and 2022 in 328 economics journals listed in the 2018 ABS Journal Guide. A fundamental challenge in this investigation is the objective assessment of title attractiveness. To overcome this, we develop a novel two-stage methodology combining human evaluation with advanced natural language processing techniques. This approach allows us to analyze title attractiveness on an unprecedented scale in economics literature. First, we conduct a stratified sampling procedure based on ABS journal rankings to select 1900 titles from our dataset. We then recruit 14 annotators, strategically chosen to represent diverse academic backgrounds and career stages, to evaluate these titles. In the second stage, we leverage the state-of-the-art deep learning model BERT (Bidirectional EncoderRepresentations from Transformers) (Devlin, Chang, Lee, and Toutanova, 2018) to learn from and synthesize the human annotations. BERT has already demonstrated its utility in various cutting-edge economics and management science research(Ash and Hansen, 2023), including sentiment analysis (Gorodnichenko, Pham, and Talavera, 2023), text classification (Zhang, Yang, Zhang, and Palmatier, 2023), and information extraction from unstructured data (Hansen, Lambert, Bloom, Davis, Sadun, and Taska, 2023). This approach ensures a comprehensive representation of title preferences across various economics subfields and levels of academic experience, while maintaining feasibility and scalability.

Our empirical analysis reveals that attractive titles significantly enhance both publication outcomes and citation counts. Specifically, papers with attractive titles are published in journals with ratings 0.120 points higher, on average, than those without such titles. Moreover, attractive titles are associated with an increase of 1.925 citations per paper. These effects remain statistically significant after controlling for a comprehensive set of author, institutional, and journal characteristics. The results are robust to more stringent specifications, including the inclusion of journal, institution, and author fixed effects, which consistently align with our baseline findings. Furthermore, our conclusions remain stable across alternative journal ranking systems.

A particularly intriguing finding emerges when we employ large language models (LLMs) such as GPT-4, ChatGPT-3.5, or LLAMA3 to assess title attractiveness, bypassing the BERT model fine-tuned on human survey data. The assessments from these LLMs show different patterns that align with their respective capabilities, while also revealing an interesting trade-off between publication success and citation impact. GPT-4's evaluations most closely mirror those of our human-trained BERT model for journal quality predictions, suggesting its stronger ability to capture expert preferences in specialized fields. However, titles identified as attractive by GPT-4 show relatively modest citation gains. Conversely, ChatGPT-3.5 demonstrates superior performance in predicting citation impact, with its identified attractive titles generating substantially higher citation counts, despite showing weaker correlation with journal quality. This trade-off potentially reflects the distinct nature of these two outcomes: journal quality assessment typically involves expert evaluation adhering to field-specific standards, while citations may be driven by broader academic appeal. These findings not only validate our baseline estimates but also suggest that different types of title attractiveness might separately optimize for expert evaluation versus general scholarly impact.

Further investigation reveals heterogeneity in the impact of attractive titles across journal tiers. For articles published in lower-tier journals (e.g., those rated one or two stars), title attractiveness shows no significant effect on citation counts. However, the impact becomes pronounced for articles in middle and upper-middle-tier journals (e.g., those rated three or A-level), with the effect being particularly strong for papers published in 4-star journals, where an attractive title is associated with up to 13.7 additional citations. Interestingly, this effect diminishes for publications in the very top journals (4* or AA-level), suggesting a non-linear relationship between journal prestige and the marginal benefit of an attractive title. This pattern aligns with intuition: while papers in lower-tier journals may struggle to attract attention regardless of title attractiveness, and top-tier papers may be widely read due to journal prestige alone, papers in middle-tier journals appear to benefit most from eye-catching titles that help them stand out in an increasingly competitive academic landscape.

Our study makes three primary contributions to the literature on academic publishing in economics and the broader field of scientometrics. First, we provide novel empirical evidence on the causal role of title attractiveness in determining publication outcomes and citation counts in economics. While previous literature has focused on correlations between observable title characteristics (e.g., length, punctuation) and academic success (Gnewuch and Wohlrabe, 2017; Bramoullé and Ductor, 2018; Guo et al., 2018), we demonstrate that title attractiveness, a subjective measure, significantly influences both publication venue prestige and subsequent citation rates, independent of these observable features. We posit that attractiveness serves as the primary channel through which various title attributes affect academic outcomes, offering a unifying explanation for previously reported effects. This mechanism, grounded in behavioral economics theories such as limited attention (Sims, 2003) and framing effects (Tversky and Kahneman, 1981), moves beyond correlational findings to propose a causal link between title characteristics and academic success. Our contribution provides a more nuanced framework for understanding the determinants of academic impact in economics (Heckman and Moktan, 2020).

Second, we introduce a novel methodological approach for investigating the impact of subjective paper characteristics on academic outcomes at scale. Traditional literature has been constrained to either large-sample studies of easily quantifiable features (Guo et al., 2018) or small-sample analyses of subjective characteristics relying on manual evaluation (Feld et al., 2024). Our methodology, leveraging advanced natural language processing techniques, bridges this gap by enabling large-scale analysis of subjective features. This innovation substantially reduces the cost and expands the scope of research into qualitative paper characteristics, opening new avenues for exploring subtle factors influencing academic success in economics and potentially other disciplines.

Finally, our study contributes to the emerging literature on the use of large language models (LLMs) in social science research (Chen, Liu, Shan, and Zhong, 2023; Horton, 2023). We find that assessments of title attractiveness derived from off-theshelf LLMs yield results consistently similar to those obtained using a BERT model fine-tuned on human annotations. This finding extends the applicability of LLMs in social sciences beyond traditional economic experiments, suggesting that these models can generate data of comparable quality to human judgments even for subjective tasks in nonexperimental settings. Our results thus broaden the potential applications of LLMs in economic and social science research methodologies, particularly for tasks involving subjective assessments that have traditionally required human evaluation.

II. DATA AND METHODS

This section outlines our empirical strategy for identifying the causal impact of title attractiveness on publication outcomes and citation rates in economics. We construct a novel, comprehensive dataset and employ an innovative methodological approach to address potential endogeneity concerns and measurement challenges.

Our dataset comprises 347,197 articles published between 2000 and 2022 in 328 economics journals listed in the 2018 ABS Academic Journal Guide. We integrate metadata from the OpenAlex database with detailed author information from the Semantic Scholar database, providing a rich set of controls for our empirical analysis.

To overcome the inherent subjectivity in measuring title attractiveness, we develop a two-stage approach. First, we conduct a stratified random sampling based on ABS journal rankings, selecting 1900 titles for human evaluation. This evaluation is carried out by a diverse panel of ten professors representing various economics subfields and economics-oriented graduate students, mitigating potential biases. Second, we leverage these expert judgments to fine-tune a BERT (Bidirectional Encoder Representations from Transformers) model, allowing for scalable and consistent assessment of title attractiveness across our entire corpus.

The subsequent sections elaborate on our data sources and sample construction (2.1), detail our novel approach to measuring title attractiveness, including both the survey design and machine learning implementation (2.2), and present descriptive statistics (2.3).

II.1 Data

Our analysis is underpinned by a comprehensive dataset of economics publications, constructed through a systematic three-stage process. We begin with the 2018 ABS Academic Journal Guide to identify relevant economics journals. Subsequently, we extract article-level metadata from the OpenAlex database for publications between 2000 and 2022. Finally, we enrich this information with detailed author data from the Semantic Scholar database. The following subsections detail each data source and our sample construction methodology.

II.1.1 ABS Journal Guide

The Academic Journal Guide (AJG), published by the Chartered Association of Business Schools (ABS), serves as a critical benchmark for assessing journal quality in business and management disciplines, including economics. This guide plays a pivotal role in academic evaluations, informing decisions on faculty hiring, promotion, and research quality assessment exercises (Walker, Fenton, Salter, and Salandra, 2019). The AJG employs a nuanced star-based rating system, categorizing journals from 1 (lowest) to 4* (highest), reflecting the relative impact and prestige within their respective fields.

For our empirical analysis, we employ the 2018 edition of the AJG. While more recent editions are available, the 2018 version provides a consistent benchmark for our analysis period (2000-2022). The choice of this particular edition is predicated on the observation that economics journal rankings have exhibited relative stability across recent editions, although this assumption merits further scrutiny in future research.

From the 2018 AJG, we extracted a comprehensive set of 336 economics journals, including their titles and International Standard Serial Numbers (ISSNs). The journal quality distribution in our sample exhibits considerable variation: 6 journals are classified as 4* (coded as 5 in our dataset to maintain ordinal consistency), 17 as 4, 67 as 3, 122 as 2, and 124 as 1. This heterogeneity in journal quality provides a rich backdrop for our investigation into the determinants of publication success.

To further refine our dataset and account for potential confounding factors, we conducted a manual review of each journal's official website. This process allowed us to categorize journals based on two additional characteristics: whether they are general field journals and whether they primarily publish theoretical work. We posit that these journal attributes may influence title construction and, consequently, impact our analysis of title attractiveness. By incorporating these manually collected variables, we enhance our ability to isolate the effect of title attractiveness on publication outcomes, controlling for journal-specific factors that might otherwise bias our estimates.

The AJG ratings serve a dual purpose in our study. Primarily, they function as a key dependent variable, offering a proxy for publication prestige in our empirical strategy. Additionally, the journal titles and ISSNs extracted from the AJG, along with our manually collected journal characteristics, form the foundation for our subsequent data collection efforts via the OpenAlex API. This comprehensive approach enables the compilation of a rich, article-level dataset spanning our study period, equipped with nuanced journal-level controls that strengthen the robustness of our analysis.

II.1.2 OpenAlex

Our dataset is constructed upon the OpenAlex database, an open-source, comprehensive index of scholarly works, authors, venues, institutions, and concepts (Priem, Piwowar, and Orr, 2022). This resource provides a wealth of bibliometric data crucial for our analysis.

We conducted a targeted query of the OpenAlex API using the 336 journal ISSNs and titles identified from the ABS Journal Guide, extracting all publications from these journals between 2000 and 2022. For each article, we collected an extensive set of metadata, including title, publication year, DOI, open access status, publisher, citation metrics (total and yearly distribution), keywords, reference count, abstract, and detailed author information (affiliation, order, and corresponding author status).

Our initial data extraction yielded 353,633 articles. However, information for 15 journals, predominantly lower-ranked (1-2 star) economics outlets, was unavailable through the OpenAlex API. This minor limitation in coverage is unlikely to significantly affect our results, given the breadth and depth of our dataset.

To ensure the integrity and relevance of our sample, we implemented a rigorous data cleaning protocol. This process entailed the removal of non-English publications to maintain linguistic consistency, the exclusion of non-research items (e.g., notes, comments, prefaces), and the elimination of articles with extremely short (¡ 6 characters) or long (¿ 150 characters) titles to mitigate potential data entry errors or non-standard entries. We also excised titles containing an excessive number of special characters and performed a de-duplication process to prevent double-counting. Post-cleaning, our final dataset comprises 325,203 unique articles, representing a robust and comprehensive sample of economics literature spanning two decades.

II.1.3 Semantic Scholar

The characteristics of authors are crucial determinants of publication outcomes and citation patterns in academic research. As articles are fundamentally the product of their creators, author attributes likely play a significant role in shaping title attractiveness, publication success, and citation counts. However, the author information available in the OpenAlex dataset is limited to names, authorship order, and institutional affiliations, necessitating the incorporation of additional data sources to comprehensively control for author-specific effects.

To address this limitation, we augment our dataset with information from the Semantic Scholar database. Launched in 2015 by the Allen Institute for Artificial Intelligence, Semantic Scholar is an open data platform designed to facilitate efficient navigation of the vast scientific literature landscape. Utilizing advanced natural language processing and machine learning techniques, it extracts and analyzes data from over 200 million papers to create the Semantic Scholar Academic Graph, a comprehensive network of papers, authors, and citations. The platform offers features such as personalized paper recommendations, detailed author pages, and APIs for accessing scholarly data, thereby accelerating scientific discovery and research (Kinney, Anastasiades, Authur, Beltagy, Bragg, Buraczynski, Cachola, Candra, Chandrasekhar, Cohan et al., 2023). In the field of economics, Semantic Scholar has been employed in recent studies to retrieve detailed bibliometric information and subject tags, enhancing the depth and breadth of economic literature analysis (Galiani, Gálvez, and Nachman, 2023).

For our study, we extracted information on 210,420 unique authors identified in our OpenAlex literature dataset. To mitigate the risk of homonymy, we implemented a rigorous matching process, ensuring that each author in our sample had at least one publication in Semantic Scholar corresponding to a publication in our OpenAlex dataset. This approach enhances the reliability of our author-level data. From Semantic Scholar, we obtained three key metrics for each author: total citation count, h-index, and publication count. These metrics serve as proxies for author reputation and productivity, factors that have been shown to significantly influence publication success and scholarly impact (Hamermesh, 2018).

Gender has been identified as a significant factor affecting academic publishing and citation patterns in economics (Bransch and Kvasnicka, 2022). However, Semantic Scholar does not provide information on author gender. To address this gap, we employed an innovative approach using ChatGPT to predict author gender based on their names. This method has been demonstrated in recent literature to achieve high accuracy, outperforming traditional pre-trained models in gender prediction tasks (Alexopoulos, Lyons, Mahetaji, Barnes, and Gutwillinger, 2023). By incorporating this gender data, we are able to control for potential gender-related effects in our analysis of title attractiveness and publication outcomes.

The integration of these comprehensive author characteristics with our OpenAlex dataset results in a rich, multidimensional dataset. This approach allows us to control for a wide array of factors that may influence title attractiveness and publication outcomes, thereby enhancing our ability to isolate the effect of title characteristics while accounting for author-specific variables.

II.2 Measurement of Title Attractiveness

In our investigation of the impact of title attractiveness on academic outcomes in economics, we face two significant methodological challenges. The first stems from the inherently subjective nature of title attractiveness, which defies simple, objective definition. This subjectivity manifests in multiple dimensions: the use of idiomatic expressions, humor, brevity, or other stylistic elements may all contribute to a title's appeal, yet the relative importance of these factors is not uniformly agreed upon. Moreover, individual perceptions of what constitutes an ”eye-catching” title can vary substantially, rendering any attempt at establishing universal criteria problematic.

This subjectivity necessitates a consensus-based approach, where title attractiveness is determined through aggregated judgments from a diverse group of evaluators across various economic subfields. However, this leads to our second challenge: the sheer scale of our dataset, comprising 325,203 article titles, makes comprehensive human evaluation prohibitively costly and time-consuming.

To address these challenges, a natural inclination might be to turn to machine learning techniques, which have been increasingly employed in economic research to analyze large-scale unstructured data. Machine learning approaches excel at extracting meaningful information from high-dimensional data (such as text) and transforming it into lower-dimensional representations (such as categories or scores)(Ash and Hansen, 2023). This capability is particularly valuable in our context, where the complex, multifaceted nature of title attractiveness needs to be distilled into a quantifiable measure.

Our task bears similarity to the concept detection problem, a typical application of machine learning in economics. Traditionally, such problems are addressed using dictionary methods within the bag-of-words model (Ash and Hansen, 2023). However, our case diverges from conventional concept detection in a crucial aspect: the subjective and multifaceted nature of title attractiveness precludes the use of a predefined dictionary.

Unlike some economic text analysis tasks where predefined dictionaries (such as financial sentiment dictionaries (Loughran and McDonald, 2011)) can be effectively employed, the concept of an ”eye-catching” title is inherently subjective and context-dependent. While it's theoretically possible to construct a dictionary for this purpose, doing so presents unique challenges. If we were to build a dictionary based on our own rules or perceptions, we would inevitably introduce our personal biases, potentially failing to capture what constitutes an eye-catching title across diverse economic subfields and for a general academic audience. To truly reflect a consensus view of title attractiveness, we would need to engage annotators from various economic disciplines to independently construct dictionaries, which would then be synthesized into a comprehensive lexicon. However, this approach would be prohibitively costly and time-consuming. Moreover, it might still fall short of capturing the nuanced and context-dependent aspects of title attractiveness that go beyond mere word choice.

Consequently, the most suitable and direct approach is to employ machine learning algorithms to predict title attractiveness based on a subset of human annotations (Ash and Hansen, 2023), rather than attempting to construct a comprehensive dictionary. This method involves selecting a representative group of annotators to evaluate a random subset of titles, generating labels that can then be used to train a machine learning model to predict the attractiveness of the remaining titles.

Conventional approaches often combine machine learning models with the bag-of-words representation for text analysis tasks. For example, Jelveh, Kogut, and Naidu (2024) use the 98479 phrases from 20029 economics papers to construct the bag-of-words model, and then use the groundtruth ideology of the authors of these papers as lables to train a model to predict the ideology of economics papers. This method typically involves creating a word count matrix from the corpus and using it to train supervised learning models (Dell, 2024). However, in the context of assessing title attractiveness, the bag-of-words approach faces two significant limitations: First, high word overlap between attractive and unattractive titles, meaning they are similar in the embedding space but different in their attractiveness (Dell, 2024). Second, it struggles to capture the nuanced use of words and their contextual relationships, which is critical in judging the title attractiveness (Ash and Hansen, 2023). Consider the following two titles:

1. "Time to include time to death? The future of health care expenditure predictions"
2. "Including time to death in health care expenditure predictions: A temporal analysis"

Both titles contain similar key words such as "time", "death", "health care", "expenditure", and "predictions". However, the first title is noticeably more attractive. A bag-of-words model, which primarily focuses on word frequencies, would struggle to capture this difference in appeal.

Second, the attractiveness of a title often stems from the synergistic effect of word combinations and their contextual usage, rather than the mere presence of individual "attractive" words. Bag-of-words models, by design, disregard word order and lose contextual information, which are crucial for understanding title attractiveness. In our examples, words like "time", "death", or "future" are not inherently eye-catching in isolation. However, their combined use and contextual arrangement in the first title creates a compelling narrative. The phrase "Time to include time to death?" juxtaposes the concept of time in two different contexts, creating an intriguing wordplay. Similarly, "The future of health care expenditure predictions" places a forward-looking spin on what could otherwise be a dry topic. These subtle linguistic features—word order, phrasing, and rhetorical devices—contribute significantly to the title's appeal but are lost in a bag-of-words representation. The model would treat ”time to include time to death” simply as individual word counts, missing the clever repetition and question structure that makes it engaging.

Given the complexity of our task, we employ BERT (Bidirectional Encoder Representations from Transformers), one of the most advanced natural language processing (NLP) models, to address our research question. Developed by Google AI in 2018, BERT represents a significant breakthrough in NLP (Devlin et al., 2018). It is a pre-trained language model based on the Transformer architecture, sharing technological foundations with widely recognized models like ChatGPT.

BERT offers several key advantages that make it particularly suitable for assessing title attractiveness. At its core, BERT utilizes a self-attention architecture, enabling dynamic consideration of relationships between words. When processing titles, BERT can capture the contextual meaning of words and the unique effects produced by different word combinations. For instance, in evaluating a title like ”Time to include time to death?”, BERT can discern the varied meanings of ”time” in different positions and appreciate the rhetorical effect of this repetition.

Furthermore, BERT's distinctive pre-training method involves randomly masking input text words and training the model to predict these masked words. For example, given the sentence ”The [MASK] brown fox jumps over the [MASK] dog,” BERT attempts to predict the masked words. This training approach mimics human language comprehension processes, endowing BERT with robust language understanding capabilities crucial for detecting subtle linguistic features in titles.

A key strength of BERT lies in its large-scale pre-training and transfer learning capabilities. BERT is pre-trained on an extensive corpus of text data, including Wikipedia and BooksCorpus, amassing a wealth of linguistic knowledge. Specifically, BERT's training data comprises 3.3 billion words from English Wikipedia and 800 million words from BooksCorpus (Devlin et al., 2018). This comprehensive pre-training enables BERT to develop a broad and deep understanding of natural language. Through transfer learning, we can adapt BERT to our specific task of assessing title attractiveness with a relatively small amount of annotated data, quickly achieving high performance (Dell, 2024).

To understand the idea of fine-tune and transfer learning [1], we can analogize it to human education process. Attempting to train a nerual networkmodel from scratch to judge whether economics titles are eye-catching would be akin to teaching a new-born baby with no prior knowledge. Even with substantial time investment, the child might

[1] More details of Bert and idea of transfer learning can be refer to Dell (2024)

struggle to understand the nuances that make an economics title attractive if they are only exposed to the examples we collected, since they have no comprehensive understanding of the language. In contrast, using a pre-trained BERT model is comparable to tasking a well-read university student, familiar with a wide range of topics including Wikipedia content, with learning to assess the attractiveness of economics titles. This ”student” already possesses rich linguistic knowledge and understanding, requiring only minimal specialized training to effectively discern eye-catching titles in economics.

The transfer learning paradigm, or fine-tuning of BERT, typically involves adding a linear classifier to the pre-trained BERT model. This approach usually focuses on training the parameters of the classifier while keeping the pre-trained BERT parameters fixed (Dell, 2024). This methodology has been widely recognized as successful and has been applied across various domains in computer science and related fields. For instance, in e-commerce, BERT has been applied to sentiment analysis of product reviews on platforms like Amazon (Sun, Qiu, Xu, and Huang, 2019). It can accurately classify reviews as positive or negative, even capturing nuanced language use such as sarcasm and context-dependent expressions. In the biomedical domain, researchers have adapted BERT for text mining of scientific literature, creating specialized versions like BioBERT (Lee, Yoon, Kim, Kim, Kim, So, and Kang, 2020). These models excel at tasks such as named entity recognition, where they can automatically identify and categorize mentions of diseases, drugs, and genes in medical texts. In the field of management science, BERT is also very popular to used in the frontier research in recent years. Kovács, Hsu, and Sharkey (2024) use BERT to predict categorization of a given book based solely on an author’s description of its content. Zhang et al. (2023) utilized BERT to extract opinions from consumer reviews, demonstrating its capability in sentiment analysis within a marketing context. All of these research find fine-tuned BERT model achieve great performance on their task and highly outperforms the existing models in their respective tasks.

Despite the relatively limited application of BERT in economic research thus far, numerous scholars recognize the significant potential of deep learning models like BERT in advancing economic studies. This potential manifests in two primary dimensions. Firstly, BERT offers a direct means to incorporate multilingual, unstructured textual data into empirical research. It can map and extract low-dimensional vectors, such as categories, topics, or sentiments, from unstructured data.(Ash and Hansen, 2023) This capability substantially enriches the data sources available for economic research, allowing for the analysis of previously untapped information. Secondly, these models are particularly well-suited for large-scale annotation and extraction tasks. By fine-tuning deep learning models on small-scale annotated datasets, researchers can efficiently annotate and extract information from extensive datasets. This approach has the potential to significantly enhance the efficiency of economic research while reducing associated costs (Dell, 2024).

Dell (2024) demonstrates the efficacy of this approach in a historical context. In their study, they manually annotated historical articles into specific topics and subsequently used this annotated data to fine-tune a BERT model. Remarkably, they found that BERT exhibited excellent transfer learning performance with a limited amount of annotated data (300-1000 samples). The model achieved over 90% accuracy across 19 different topic classifications, illustrating the feasibility of using BERT for large-scale data annotation and extraction in economic research.

In recent years, BERT has been employed in several pioneering and influential studies in economics. To the best of our knowledge, the earliest application of BERT in economics is found in Shapiro, Sudhof, and Wilson (2022). They trained a BERT model to predict the sentiment of news articles. However, due to the length of their training texts, limited labeled data (800 news articles), and a broad categorization scale (Very Negative (1) - Very Positive (5)), the performance of BERT did not surpass their innovative and orginal lexicon-based method. In pursuit of transparency and ease of interpretation, they ultimately relied on their lexicon-based approach for subsequent empirical analyses.

A notable and successful application of BERT in economics is presented by Gorodnichenko et al. (2023). Their study investigates the impact of Federal Open Market Committee (FOMC) meeting tones on financial markets. To independently identify the tone of FOMC meetings, they needed to control for the sentiment of policy messages. Their methodology involved segmenting FOMC statements into sentences and employing research assistants to label these sentences on a scale from very hawkish (-10) to very dovish (10). The final score for each FOMC statement was derived from the average of these sentence scores. Subsequently, they utilized word embeddings from BERT as input to train a neural network, effectively implementing a fine-tuning approach[2]. Their fine-tuned BERT model achieved over 80% accuracy in predicting hawkish, neutral, and dovish stances of FOMC statements.

In more recent research, BERT has been applied to large-scale classification tasks, such as determining whether job postings are remote-work friendly. Hansen et al. (2023), in their study of the pandemic-catalyzed shift to remote work, fine-tuned a BERT model to assess whether 250 million job vacancy postings were amenable to remote work. Their methodology involved dividing job postings into sequences and employing three annotators to label these sequences as either "Remote-work-friendly" or

[2]This approach is equivalent to fine-tuning, as it involves adding a new classifier to the base BERT model and training this classifier to achieve transfer learning (Dell, 2024)

"Not-remote-work-friendly". Utilizing 30,000 human classifications for training, their model achieved 99% accuracy in identifying job postings advertising hybrid or fully remote work, significantly outperforming both dictionary methods and other machine learning approaches.

Given the complexity of our research question, the advanced capabilities of BERT, and its demonstrated success in various economic studies, we propose a two-stage strategy to leverage BERT for assessing title attractiveness. This approach allows us to balance the need for accurate classification with cost considerations.

We begin by selecting a representative group of annotators from the economics field to label a subset of our data. To ensure a diverse perspective, we recruited seven economics professors and seven economics students from various sub-disciplines, who evaluated the attractiveness of titles from a stratified sample of 1900 papers. This diverse pool of annotators helps to mitigate potential biases and ensures that our trained BERT model captures a broad consensus on what constitutes an attractive title in economics.

The following subsections detail our annotation experiment (2.2.1) and the BERT model fine-tuning process (2.2.2).

II.2.1 Survey Design and Implementation

The initial phase of our title attractiveness measurement involved a carefully designed annotation experiment. We solicited evaluations from both economics scholars and students on a subset of paper titles, creating a training dataset that captures the collective preferences of the economics academic community. Employing a stratified sampling technique based on the ABS journal star ratings, we selected 1900 papers from our total sample, ensuring proportional representation across journal quality tiers and mitigating potential biases in our training data. The selected titles were randomized and divided into subsets for individual annotator evaluation.

To isolate the effect of title attractiveness, we presented annotators with only the paper titles, withholding all other metadata. This design choice minimizes the influence of confounding factors such as author reputation or journal prestige on the attractiveness assessment. Annotators were prompted with the following instruction:

> "Please evaluate the following economics paper titles and determine if they are eye-catching. Label with 0 or 1. If you find the title very eye-catching and it makes you want to read the paper, label with 1; otherwise, label with 0."

Our annotator pool comprised both professors and students from diverse subfields of economics, addressing two primary concerns. First, it mitigates potential domain-specific biases in title attractiveness judgments. During pilot discussions, several annotators noted that their perceptions of title appeal were influenced by their research interests. By diversifying our annotator pool across economics subfields, we aim to capture a more representative, field-agnostic measure of title attractiveness. This approach helps to address potential endogeneity concerns, such as the possibility that papers from certain subfields might be more likely to be published in high-ranking journals or accrue more citations due to field-specific factors rather than title attractiveness.[3]

Second, the inclusion of both professors and students serves to balance expertise-driven and naive assessments of title attractiveness. While professors, given their extensive reading experience, might inadvertently associate "eye-catching" titles with those typical of top-tier publications or highly-cited papers, students with less exposure to the literature may provide more unbiased assessments based purely on the title's appeal. This combination helps to mitigate potential reverse causality issues where title assessments could be influenced by recognition of high-status papers rather than intrinsic attractiveness.

Our annotator pool consisted of seven professors, ranging from assistant professors to full professors, and seven students. The student group included M.Phil. students, current Ph.D. candidates, full-time research assistants, and master's students preparing for Ph.D. studies. For the latter category, we ensured that each had at least one working paper and substantial research experience to guarantee a basic understanding of academic conventions.

We conducted our data collection in two phases. The first phase involved one professor and four students, each annotating 200 titles. This initial dataset was used to validate the BERT model's ability to learn human preferences. The second phase expanded to six professors and three students, each evaluating 100 titles. The data from both phases were combined to train the final BERT model, ensuring a comprehensive learning of collective preferences across different levels of academic experience and subfields.

Table 1 show us the detail of our annotator pool, including their type (student or professor) and their research field.

In total, we collected 1,900 title annotations from our diverse pool of annotators. Among these annotations, 266 titles were identified as "eye-catching," suggesting that our annotators maintained relatively high standards in their evaluations. Interestingly, we

[3] While our experimental design aims to mitigate these concerns, we further control for journal categories in our baseline regressions and JEL codes in our robustness checks to address any remaining field-specific effects.

TABLE 1: Research Fields of Students and Professors

Number	Type	Research Field
1	Student	Political Economy; Environment Economics
2	Student	Game Theory; Political Economy
3	Student	Labour Economics
4	Student	Health Economics; Labour Economics
5	Student	Urban Economics
6	Student	International Economics
7	Student	Environment Economics
8	Professor	Health Economics; Labour Economics
9	Professor	Health Economics; Labour Economics; Development Economics
10	Professor	Labour Economics, Economics of Education
11	Professor	International Trade, Industrial Organization, Spatial Economics
12	Professor	Development Economics, Environment Economics
13	Professor	Industrial Organization, Quantitative Marketing, Health Economics
14	Professor	Energy Economics; Environment Economics

observed substantial heterogeneity in individual annotation patterns. The most conservative evaluation came from a professor who identified only 3 titles in his/her 100-title sample as eye-catching, while the most liberal assessment came from a student who classified 38 titles in his/her sample as eye-catching. This variation in annotation patterns across academic ranks and experience levels underscores the subjective nature of title attractiveness and validates our strategy of incorporating diverse perspectives in our training data.

The observed heterogeneity in annotations also provides valuable insights for our subsequent machine learning approach. Rather than indicating inconsistency in our data, these variations suggest that title attractiveness is perceived differently across different levels of academic experience and expertise. This diversity in our training data is particularly valuable for our BERT model, as it allows the model to learn a more comprehensive representation of title attractiveness that captures various perspectives within the economics academic community.

II.2.2 BERT Model Fine-tuning

After obtaining the annotated samples, we proceeded to fine-tune the BERT model. We employed the bert-base model from Hugging Face[4], addressing two critical imbalance issues during the fine-tuning process.

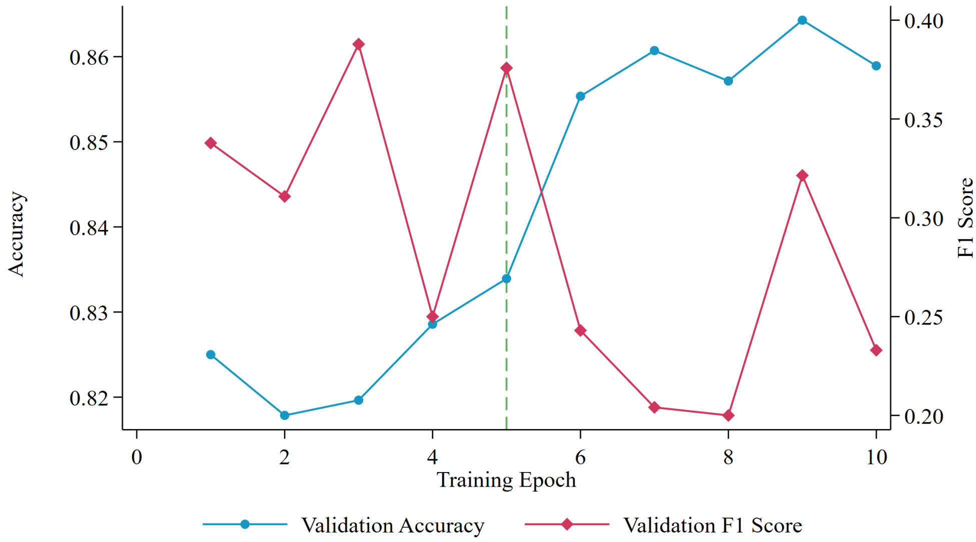
[4]https://huggingface.co/google-bert/bert-base-uncased

The first challenge was the imbalance between positive and negative samples. Since our dataset predominantly consisted of negative samples (non-eye-catching titles), the model might be predisposed to negative classifications. To address this, we implemented a weighted cross-entropy loss function, where the contributions of positive and negative samples to the loss function were weighted according to their respective proportions in the training data. This approach ensures that misclassifying a positive sample as negative incurs a higher penalty than misclassifying a negative sample as positive, thereby counteracting the inherent bias in the sample distribution.

The second challenge stemmed from the uneven distribution of annotations across annotators. Of our 1,900 samples, 1,000 were contributed by five annotators, while the remaining 900 came from nine annotators. This disparity could potentially skew the model's learned preferences toward those of the more prolific annotators. To mitigate this bias, we employed an oversampling strategy for the samples from the nine annotators with fewer annotations, effectively doubling their 900 samples to achieve a more balanced representation. This approach simulates a scenario where each of these nine annotators had provided 200 samples, comparable to the five more prolific annotators.

Following these adjustments, our final training dataset comprised 2,800 samples. We allocated 20% (560 samples) for validation and used the remaining samples for training. The model was trained with a batch size of 16 for 10 epochs. Importantly, we only considered complete epochs in our evaluation, as fractional epochs could potentially introduce annotator-specific biases into the model's learning process. In evaluating the model's performance, we considered both accuracy and F1 score, with particular emphasis on the latter as it provides a more comprehensive measure of performance by balancing precision and recall. After comparing performance across epochs, we selected the checkpoint from the fifth epoch as our final model, which achieved an accuracy of 0.86 and an F1 score of 0.375.

FIGURE 1: Training Metrics on Validation Sets



A pertinent question arises: Is achieving an accuracy of 86% and an 0.375 F1 score satisfactory? We contend that these results are indeed quite satisfactory, and to explain why, we must first consider what BERT is actually learning. BERT essentially learns from the annotators' data, capturing their preferences for classification tasks. For instance, in determining whether a text is dovish or hawkish, BERT learns the annotators' criteria for such judgments. While there may be consensus among annotators for strongly dovish or hawkish texts, disagreements often arise for more neutral content, stemming from individual annotator preferences.

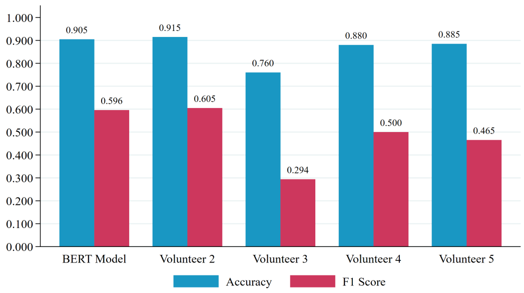
In tasks that blend objectivity with subjectivity, researchers sometimes worry that annotator biases might lead to skewed results for relatively objective annotations, potentially introducing endogeneity issues (Dell, 2024). To mitigate this, some studies use two annotators for the same data, training the model on their consensus. For such highly objective tasks, validation set accuracy must be exceptionally high, as the training data represents a strong consensus that approaches an objective standard.

However, our task of determining whether a title is eye-catching is inherently more subjective. The lack of a strong objective standard means that using annotations from different individuals may not necessarily yield high accuracy on the validation set. This is because there isn't a highly consistent consensus on what constitutes an eye-catching title. Even a well-trained model cannot perfectly align with all 20 annotators' preferences simultaneously, as some preferences may be non-overlapping or contradictory.

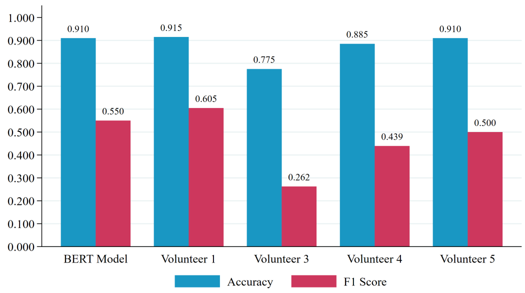
Our goal, therefore, is for BERT to learn a composite preference from these 20 individuals, or more precisely, to capture the areas of near-consensus among them. BERT achieves this through a training process analogous to a firm's profit maximization strategy in a market with heterogeneous consumer preferences. Just as a firm would position its product to appeal to mainstream consumer preferences to maximize profits, BERT adjusts its parameters to capture the most common patterns in the annotated data. This approach is reminiscent of Hotelling's location model in spatial economics (Hotbllino, 1929), where firms tend to locate at the median of consumer preferences to maximize market share. By optimizing across all annotators, BERT effectively "locates" its predictions at the center of the distribution of human judgments. This approach allows the model to capture areas of agreement while moderating highly subjective or contentious judgments, resulting in a learned representation that reflects the collective wisdom of the annotators.

FIGURE 2: Bert Performance Compared with other human annotators

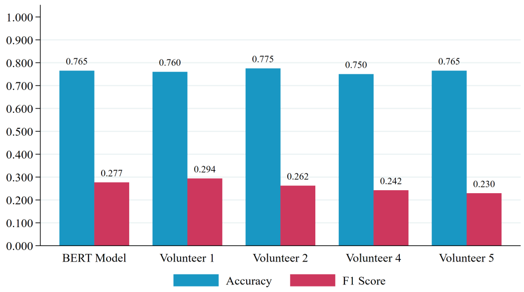
(A) Compared with Volunteer 1



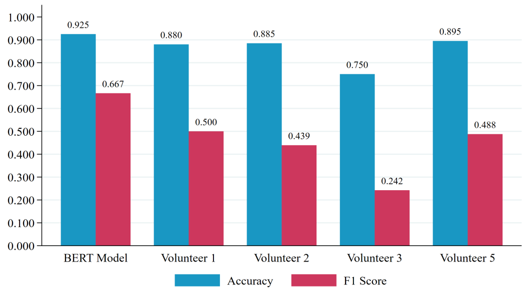
(B) Compared with Volunteer 2



(C) Compared with Volunteer 3



(D) Compared with Volunteer 4



(E) Compared with Volunteer 5

To validate our hypothesis that fine-tuned BERT effectively synthesizes annotators' preferences and to demonstrate the heterogeneity in human judgments of eye-catching titles, we conducted an additional experiment using the first batch of annotators.

First, we trained a BERT model using only the annotations from the initial five annotators. This model achieved notably higher performance metrics - an accuracy of 88% and an F1 score of 0.52 on the validation set - compared to our final model trained on all 14 annotators' data (86% accuracy and 0.375 F1 score). This performance difference reflects the greater consistency within the smaller annotator group, whereas the inclusion of additional annotators introduced more diverse preferences, leading to a more challenging learning task.

To further investigate the model's ability to capture collective preferences, we conducted a comparative analysis where these five annotators and our fine-tuned BERT model evaluated the same set of 200 paper titles. In Figure 1, we present pairwise comparisons of accuracy and F1 scores between each human annotator (used as reference) and both other human annotators and the BERT model.

Two key findings emerge from this analysis. First, while human annotators show some consistency in their judgments, there is considerable variation in their preferences. The pairwise comparisons between human annotators typically yield F1 scores around 0.5 and accuracy scores around 0.85, indicating substantial individual differences in what constitutes an eye-catching title.

Second, and more importantly, our fine-tuned BERT model demonstrates remarkable consistency with human judgments across all comparisons. When compared against annotators 1, 2, and 3, the BERT model achieves the second-highest agreement scores among all pairwise comparisons. Moreover, it shows the highest agreement with annotators 4 and 5. Overall, the BERT model exhibits greater average agreement with individual annotators than any single human annotator achieves with their peers, suggesting that it has successfully learned to synthesize a composite preference that better represents the collective judgment of the group.

This analysis provides compelling evidence that our fine-tuned BERT model effectively captures and synthesizes diverse human preferences regarding title attractiveness, rather than simply mimicking any individual annotator's judgment. The model's ability to achieve consistently high agreement scores across different reference annotators suggests it has learned a balanced representation of what constitutes an eye-catching title in economics research.

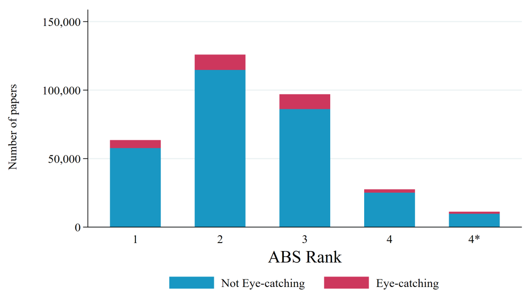
II.3 Descriptive Statistics of Title Attractiveness

In this section, we present summary statistics on our sample and examine the distribution of eye-catching papers across different dimensions.

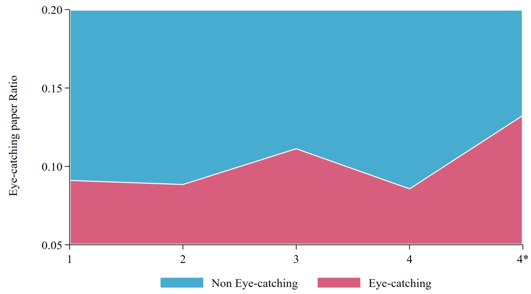
FIGURE 3: Papers Distribution Among Different Journal Quality

(A) ABS Journal Guide

Paper Numbers

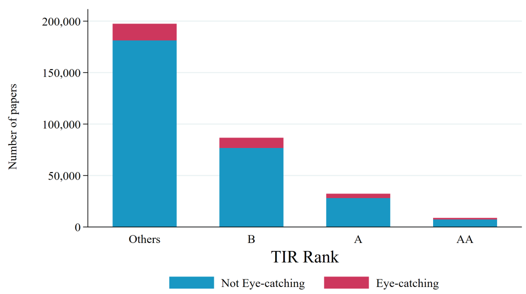


Eye Catching Papers Ratio



(B) Tinbergen Institute Ranking

Paper Numbers



Eye Catching Papers Ratio

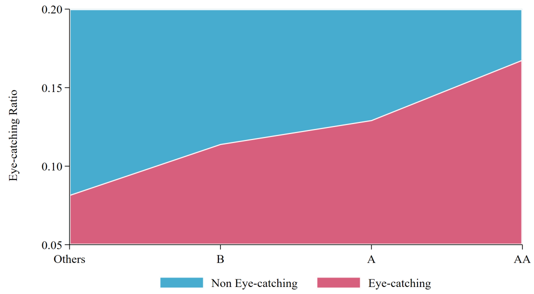
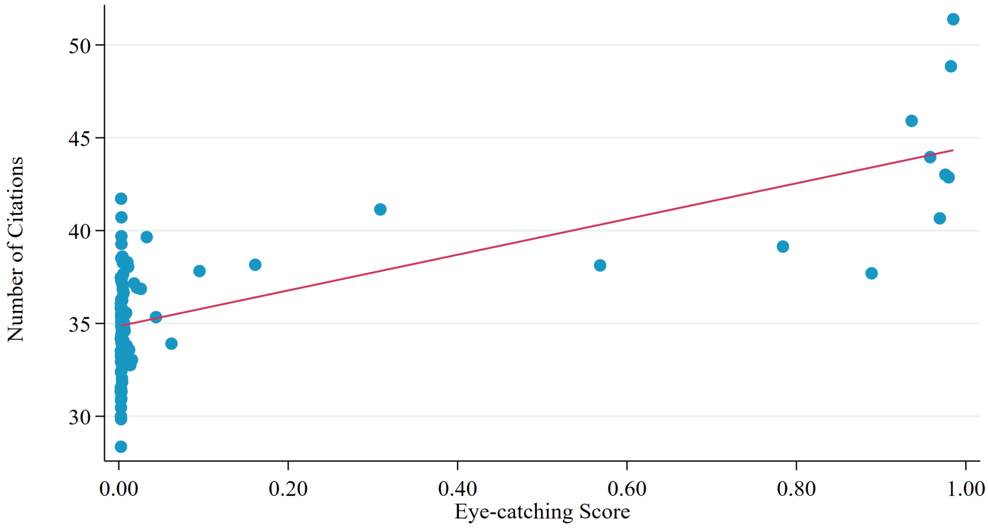


Figure 3 presents the distribution of eye-catching papers across different journal quality tiers. Panel A shows this distribution according to the ABS Journal Guide rankings, while Panel B uses the Tinbergen Institute rankings. Two key patterns emerge from this analysis. First, our sample exhibits substantial concentration in middle-tier journals. In the ABS rankings, papers published in 2- and 3-star journals constitute the majority of our sample. Similarly, in the Tinbergen Institute rankings, approximately two-thirds of the papers fall into the ”others” category. While the absolute number of eye-catching papers is also highest in these tiers, this primarily reflects the larger base number of publications rather than a higher propensity for attractive titles. Second, we observe that higher-ranked journals tend to publish a larger proportion of papers with eye-catching titles. Specifically, in the ABS rankings, while only about 8% of papers in 1-star journals have eye-catching titles, this proportion increases to approximately 14% for 4* journals. This pattern is similarly reflected in the Tinbergen Institute rankings, where A and AA-ranked journals show a notably higher proportion of eye-catching titles compared to B-ranked journals.

FIGURE 4: Eye catching vs Citations

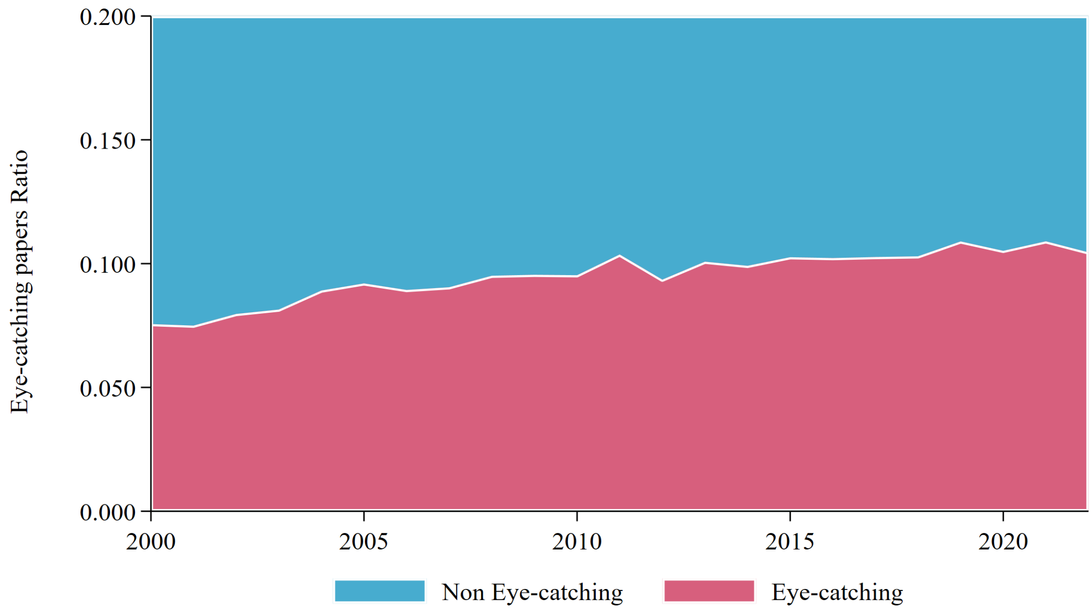


Note: Each point represents 1% of observations

Figure 4 illustrates the relationship between title attractiveness and citation counts. Here, title attractiveness is measured using the logits from our fine-tuned BERT model's hidden layer output, which can be interpreted as the model's estimated probability of a title being eye-catching (detailed discussion of this measure follows in subsequent sections). The figure reveals a clear positive association between eye-catching titles and citation impact. Papers with eye-catching titles consistently receive more citations across all publication years in our sample.

FIGURE 5: Year Trend on Eye Catching Papers

(A) Paper Numbers Contributed by Eye Catching Papers



(B) Citations Contributed by Eye Catching Papers

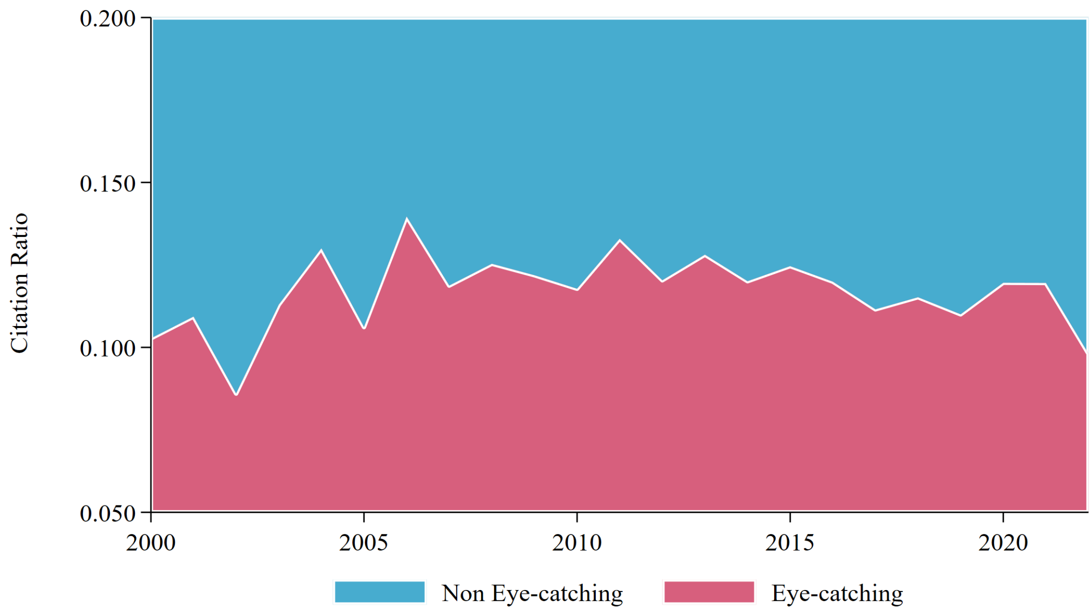


Figure 5 depicts temporal trends in eye-catching papers' contribution to the economics literature. Panel A shows the evolution in the number of eye-catching papers over time, while Panel B presents their contribution to total citations. Several interesting patterns emerge. First, we observe a steady increase in the proportion of eye-catching titles over our sample period, suggesting a growing awareness among economists of the importance of title crafting. This trend is particularly pronounced after 2010, potentially reflecting increased competition for attention in academic publishing. Second, Panel B reveals that papers with eye-catching titles account for a disproportionate share of total citations, and this disparity has grown over time. Notably, while eye-catching papers consistently comprise approximately 10% of total publications throughout most of the sample period, their contribution to total citations consistently exceeds this proportion. This suggests that eye-catching papers generate higher citation impacts than their non-eye-catching counterparts.

III. Empirical Strategy

To investigate the impact of eye-catching titles on publication outcomes and citation counts in economics, we employ two main empirical specifications. Our approach aims to address potential endogeneity concerns while isolating the causal effect of title attractiveness. We begin by examining how eye-catching titles influence the ranking of economics journals. Our first model is specified as follows:

$$\begin{aligned} \text{ABS}_i = \alpha &+ \beta_1 \text{Attractiveness}_i + \beta_2 \text{Journal}_i + \beta_3 \text{CorrAuthor}_i \\ &+ \beta_4 \text{Team}_i + \beta_5 \text{Affil}_i + X_i + \lambda_t + \epsilon_i \end{aligned} \tag{1}$$

where ABSi represents the ABS star rating of the journal in which paper i is published. Our key explanatory variable, Attractiveness$i$, is a binary indicator of whether the paper's title is considered attractive, as determined by our BERT model fine-tuned on human-annotated data.

In addressing endogeneity concerns, we focus on two critical issues. First, we consider the paper's field of study. Theoretical papers, for instance, may be less likely to have attractive titles due to their use of specialized terminology, while simultaneously being more likely to be published in highly-ranked journals. This could introduce a spurious correlation between title attractiveness and journal ranking. To mitigate this issue, we include Journal$_i$, a vector of two dummy variables indicating whether the journal is theory-focused or general-interest, based on our manual classification of 321 journals.

The second major endogeneity concern relates to author quality. More capable authors might be more likely to craft attractive titles and produce higher-quality papers, leading to better publication outcomes. This could result in a positive relationship between title attractiveness and journal ranking that is driven by author quality rather than the title itself. To address this, we incorporate three key controls: CorrAuthor$i$, the total citation count of the corresponding author from the Semantic Scholar dataset; Team$i$, the average citation count of all co-authors; and Affil_i, a vector of dummy variables indicating the corresponding author's institutional ranking.

For the institutional ranking, we utilize data from the Research Papers in Economics (RePEc) database, which provides a comprehensive ranking of economic institutions. RePEc categorizes institutions into percentile ranks, from the top 1% to the top 10%. We match the corresponding author's affiliation with these rankings and create ten dummy variables (Affil_i) representing whether the institution falls into the top 1%, top 2%, and so on up to the top 10%. Institutions not ranked in the top 10% serve as our baseline group. This granular approach allows us to capture the potential impact of institutional prestige on publication outcomes and citation counts, which may correlate with both title attractiveness and our dependent variables.

We also include a vector of additional controls, X_i, which encompasses factors that may influence publication outcomes: open access status, reference count, paper length, number of authors, and the number of female authors. These variables help to account for various aspects of paper quality and characteristics that might affect publication success. Year fixed effects, λ_t, are included to control for any time trends in publication patterns.

Our second specification examines the relationship between title attractiveness and citation counts:

$$\begin{aligned} \text{Citation}_i = \alpha &+ \beta_1 \text{Attractiveness}_i + \beta_2 \text{CorrAuthor}_i + \beta_3 \text{Team}_i + \beta_4 \text{Affil}_i \\ &+ X_i + \gamma_j + \lambda_t + \epsilon_i \end{aligned} \tag{2}$$

Here, Citation$i$ represents the total citation count of paper i at the time of data collection. The explanatory variables largely mirror those in the publication model, reflecting the similarity in factors influencing both publication outcomes and citation counts. However, a key distinction in this specification is the inclusion of journal fixed effects, γj. This allows us to compare citation counts of papers with varying title attractiveness within the same journal, effectively controlling for journal-specific factors that might influence citation patterns. Consequently, we omit the journal-level controls (Journal_i) used in the publication model.

IV. Empirical Results

TABLE 2: Baseline Estimation

	Journal Quality				Citation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eye Catching	0.082***	0.087***	0.113***	0.120***	8.227***	10.335***	4.813***	5.068***	1.925**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.793)	(0.777)	(0.747)	(0.766)	(0.767)
Correspond Citations			0.000**	-0.000			0.002***	0.001***	0.001***
			(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
Team Citations			0.000***	0.000***			0.005***	0.005***	0.004***
			(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
Theoritical			0.559***	0.527***			-4.749***	-3.144***	
			(0.004)	(0.004)			(0.668)	(0.679)	
General			0.187***	0.203***			7.887***	11.582***	
			(0.005)	(0.004)			(0.556)	(0.589)	
Year FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Journal FE	No	No	No	No	No	No	No	No	Yes
Affiliation Controls	No	No	Yes	Yes	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	No	No	No	Yes	Yes
Observations	325203	325203	301121	285803	325203	325203	301121	285803	286792

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

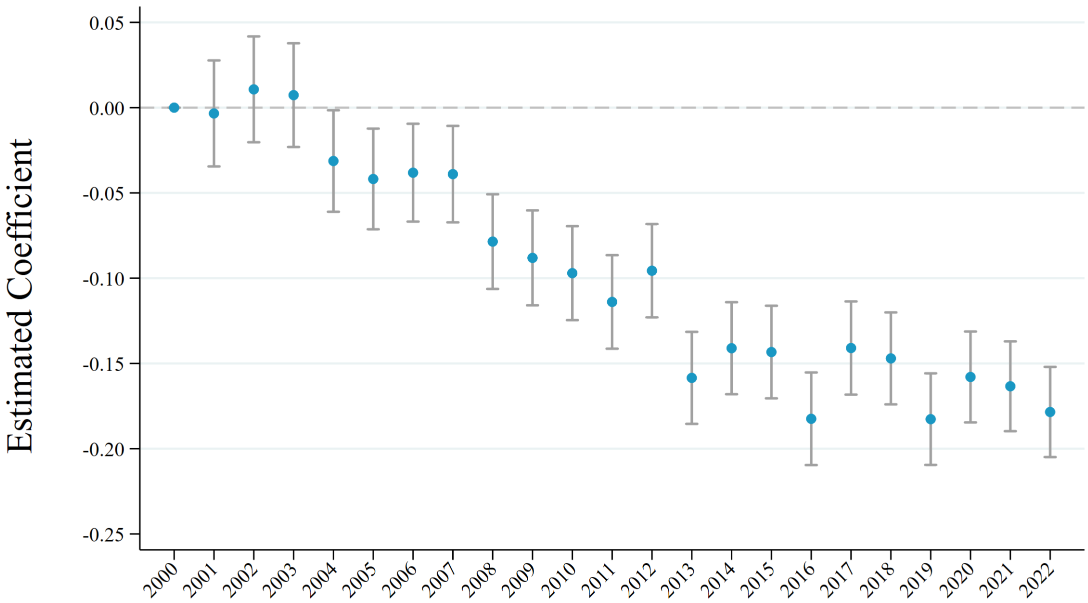
Table 2 presents the baseline estimation results for the impact of title attractiveness on publication prospects and citations. Columns (1) to (4) primarily show the impact on journal quality. In Column (1), the coefficient for our measure of title attractiveness is positive and significant at the 1% level, indicating that papers with more attractive titles are published in journals ranked 0.082 stars higher in the ABS Journal Guide compared to papers with less attractive titles. In Column (2), we control for year fixed effects to account for temporal variation in journal quality, and the coefficient remains stable. From Fig.6a, we observe that the average year fixed effect is decreasing, suggesting that, over time, the average paper's publication prospects have declined, likely due to increased competition in the field of economics. In Column (3), we add key control variables, including "Corresponding Citations", "Team Citations" and "Affiliation Controls" (to address potential endogeneity related to author quality), as well as indicators for whether the journal is "Theoretical" or "General Interest" (to account for differences in journal type). The coefficient for title attractiveness increases significantly and remains significant at the 1% level, while all key control variables also exhibit positive and significant effects on journal quality. This aligns with our expectations, as corresponding authors and author teams with strong credentials are more likely to be published in higher-quality journals. The positive coefficients for ”Theoretical” and ”General Interest” suggest that these types of journals generally have higher ABS Journal Guide rankings. In Column (4), we include additional controls, such as whether the paper is open access, the number of references, paper length, number of authors, and number of female authors. We observe a slight increase in the coefficient for title attractiveness, which remains stable and significant.

For the citation outcomes (Columns 5-9), the results similarly highlight the significant role of title attractiveness. In Column (5), the coefficient for our measure of title attractiveness is 8.227, which is positive and significant at the 1% level, indicating that papers with more attractive titles receive, on average, 8.227 more citations. This effect is further amplified in Column (6) after controlling for year fixed effects [5]. However, once key control variables are included in Column (7), the coefficient for title attractiveness decreases substantially, while remaining significant at the 1% level. This suggests that the effect of title attractiveness on citations may be partially explained by author quality and journal type. The coefficients for ”Theoretical” and ”General Interest” are also consistent with expectations: theoretical papers tend to receive fewer citations, potentially due to their complexity, while general interest papers attract more citations due to their broader appeal across disciplines. After adding additional control variables in Column (8), the coefficient for title attractiveness remains stable. In Column (9), we include journal fixed effects, which is a crucial control as it allows us to compare title attractiveness within the same journal, thereby accounting for journal-specific effects on paper quality and field of study. Although the coefficient decreases significantly to 1.925, it remains significant at the 5% level, suggesting that a paper with higher title attractiveness receives, on average, 1.925 more citations than a comparable paper with lower title attractiveness in the same journal.

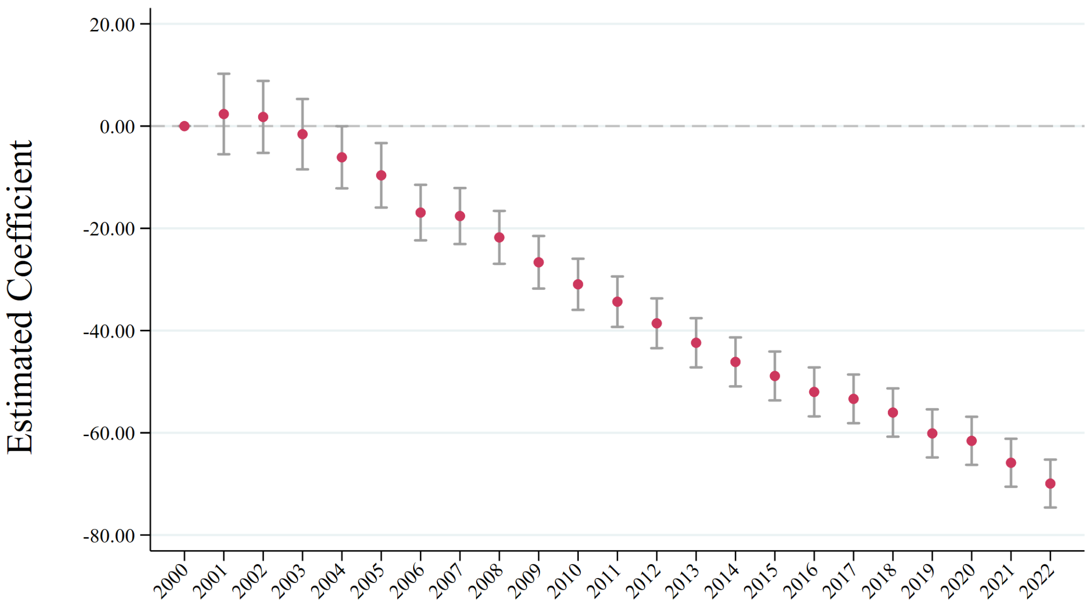
[5] From Fig.6b, we can see that average citations decrease over time, likely because newer publications receive fewer citations compared to older ones.

FIGURE 6: Year Fixed Effect Trend

(A) ABS Journal Guide



(B) Citations



V. Robustness

V.0.1 Journal Quality

Table 3: Robustness Estimation (Journal Quality)

	RePEc Rank	AA	A	B
	(1)	(2)	(3)	(4)
Eye Catching	33.412***	0.012***	0.032***	0.016***
	(2.110)	(0.001)	(0.002)	(0.003)
Corr. Citations	0.002***	0.000**	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.004***	0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	7.163***	0.018***	0.061***	-0.212***
	(1.514)	(0.001)	(0.002)	(0.002)
General	41.930***	0.091***	0.006***	0.045***
	(1.689)	(0.001)	(0.001)	(0.002)
Year FE	Yes	Yes	Yes	Yes
Affiliation Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	277747	285803	285803	285803

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In our robustness checks, we first examine the stability of journal ratings. While the ABS Journal Guide provides ratings for a wide range of economics journals, it is primarily a business school rating system, which may not be fully accurate for certain economics journals. Additionally, it includes journals from related fields like statistics, potentially introducing biases into our analysis. To address these concerns, we use two alternative journal rating metrics.

The first metric is the RePEc (Research Papers in Economics) ranking, an extensive database that ranks economics journals based on their impact over all years. The RePEc system is more closely aligned with the economics discipline's evaluation framework. Since RePEc's rankings are in ascending order (i.e., rank 1 is the highest), there may be an issue of reversed interpretability in our regression results—higher title attractiveness might appear to correlate with lower rank numbers. To ensure consistency, we reverse the ranking scale, assigning the highest-ranked journal the largest numerical value and the lowest-ranked journal the smallest. This adjustment ensures that higher journal ratings correspond to larger numerical values, making interpretation more intuitive. Table 3 presents the results of these robustness checks. In Column (1), we observe that title attractiveness still has a significant positive impact on publication outcomes. Specifically, papers with higher title attractiveness are associated with an average improvement of 33 ranks compared to papers with lower attractiveness.

The second robustness check uses the Tinbergen Institute Journal Ranking, a well-established and authoritative metric in economics. This ranking classifies journals into three categories: AA, A, and B. We treat these categories as binary variables in our regressions. Columns (2) to (4) of Table 3 show the results, indicating that title attractiveness continues to have a positive and significant effect on publication success across all categories.

Interestingly, the effect is most pronounced for A-level journals, where higher title attractiveness increases the probability of publication by 3.2%. This result aligns with our expectations. For AA-level journals, submission quality is already exceptionally high, and the review process is highly rigorous. Thus, the influence of title attractiveness is limited, as the merit of the paper is well established through extensive peer review. For B-level journals, the primary determinant is the intrinsic quality of the paper. Given the wide range of submissions, an attractive title does not necessarily indicate higher quality, which limits its impact on publication success. However, for A-level journals, most submissions achieve a high standard of quality, though not to the near-flawless level required for AA journals. In this context, the initial impression provided by an attractive title can be crucial, shaping the editorial and reviewers' perception and ultimately tipping the scales in favor of acceptance. Therefore, the effect of title attractiveness is most significant for A-level journals.

V.1 Institutions

TABLE 4: Robustness Estimation (Institution)

	Journal Quality					Citation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Eye Catch.	0.124***	0.115***	0.110***	0.114***	0.061***	1.540*	1.396*	1.511*	1.520*	1.644**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.798)	(0.798)	(0.797)	(0.798)	(0.837)
Corr. Cit.	0.000	0.000	0.000	0.000	-0.000***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Team Cit.	0.000***	0.000***	0.000***	0.000***	0.000***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Theory	0.479***	0.487***	0.483***	0.471***	0.414***					
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)					
General	0.192***	0.167***	0.178***	0.181***	0.200***					
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)					
Inst. (Pub)	0.000***					-0.001*				
	(0.000)					(0.001)				
Inst. (Cite)		0.000***					0.000***			
		(0.000)					(0.000)			
Inst. (AA)			0.001***					0.000		
			(0.000)					(0.001)		
Inst. (ABS)				0.000***					-0.000	
				(0.000)					(0.000)	
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Inst. FE	No	No	No	No	Yes	No	No	No	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	251976	251976	251976	251976	245696	252933	252933	252933	252933	246617

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In this section, we conduct a series of robustness checks to address potential concerns regarding the impact of authors' institutional affiliations on our main results. Our baseline regression incorporated RePEc's institutional rankings (top 1% - top 10%) to control for the influence of authors' institutions on title attractiveness. The underlying assumption is that higher-quality researchers are more likely to be affiliated with prestigious institutions, and individual ability may directly affect one's capacity to craft attractive titles. This relationship could potentially bias our identification strategy. To address this issue, we use institutional rankings as a proxy for individual ability, complementing the controls for corresponding author citations and team citations in our model. This approach aims to capture author-related ability characteristics that may have been omitted from our initial specification.

While RePEc's rankings are widely recognized, they cover only 1,247 institutions, whereas our dataset encompasses 10,075 unique institutions. In our baseline regression, we categorized all institutions beyond the top 1,247 as "other", potentially overlooking meaningful variation in author quality among these institutions. Given the limitations of publicly available comprehensive institution rankings, we developed alternative measures using our dataset, which covers publications from 2000 to 2022 across all 10,075 institutions. We employed four distinct metrics to rank institutions: (1) Total publication count (2) Total citation count (3) Cumulative ABS journal star ratings (4) Number of publications in journals listed in the Tinbergen Institute Journal Ranking (AA, A, B categories)

The ranking score for each institution was calculated using the following formula:

$$\text{Rank Score}_i = \sum_{j=1}^{n} \text{score}_j \times \text{Number of authors}_{ij} \tag{3}$$

where Rank Score$_i$ is the total ranking score for institution i, j indexes the papers in our sample, score$_j$ is the value assigned to paper j based on the chosen metric, and Number of authors$_{ij}$ is the number of authors from institution i on paper j. We then used the highest institutional score among all authors of a paper as our control variable.

Table 4 presents the results of our robustness checks. Columns (1)-(4) show the results using our newly developed institutional scores as control variables. We find that the coefficient on our measure of title attractiveness remains positive and significant at the 1% level across all specifications. Moreover, the coefficients on the institutional ranking measures are consistently positive and significant, confirming that institutional prestige indeed plays a positive role in publication outcomes. Column (5) employs the most stringent control by including fixed effects for the corresponding author's institution. Although the coefficient on title attractiveness decreases in magnitude, our results remain robust and statistically significant.

Columns (6)-(10) examine the robustness of the relationship between title attractiveness and citations. In Columns (6)-(9), we control for the aforementioned institutional scores. While the coefficient on title attractiveness decreases, it remains significant at the 10% level. Interestingly, the impact of institutional scores on citations is mixed and largely insignificant. This finding is reasonable, as citation behavior is typically more influenced by the journal in which a paper is published rather than the authors' institutional affiliations. Finally, Column (10) incorporates institutional fixed effects for citations as well. Our results continue to hold, demonstrating the robustness of our findings to this most stringent specification.

V.2 Authors

TABLE 5: Robustness Estimation (Corresponding Author)

	Journal Quality			Citation		
	(1)	(2)	(3)	(4)	(5)	(6)
Eye Catching	0.120***	0.119***	0.025***	1.952**	1.975***	1.035
	(0.006)	(0.006)	(0.007)	(0.767)	(0.765)	(1.287)
Team Citations	0.000***	0.000***	0.000***	0.006***	0.006***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.528***	0.520***	0.120***			
	(0.004)	(0.004)	(0.009)			
General	0.203***	0.199***	0.193***			
	(0.004)	(0.004)	(0.006)			
Corresponding Hidex	0.001***			-0.206***		
	(0.000)			(0.042)		
Corresponding Paper		-0.001***			-0.082***	
		(0.000)			(0.005)	
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	No	No	Yes	Yes	Yes
Author FE	No	No	Yes	No	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	285803	285803	299847	286792	286792	300897

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In this section, we further examine the robustness of our results with respect to the characteristics of the corresponding authors. Specifically, we replace the citation count of the corresponding author with their H-index and the number of papers they have published. These data are sourced from the Semantic Scholar database. Additionally, we adopt the most stringent approach to control for the fixed effects of corresponding authors, considering individual effects for 113,632 authors in the sample. Table 5 presents the results of our robustness checks.

Columns (1) to (3) display the robustness of the effect of title attractiveness on publication outcomes. In Columns (1) and (2), we find that even after modifying the measure of the corresponding author's ability, the coefficient of our title attractiveness variable remains robust and significant. Interestingly, when using the number of papers published by the corresponding author as a measure of their ability, the coefficient turns negative. This outcome aligns with intuition in the economics field, where scholars typically prioritize the quality of journal publications over sheer quantity. Therefore, a high paper count may not necessarily convey a positive signal about an author's research impact or ability to publish in high-quality journals.

In Column (3), where we control for extensive corresponding author fixed effects, the coefficient for title attractiveness becomes considerably smaller, but it remains significant at the 5% level. This suggests that while there is a strong association between title attractiveness and the characteristics of individual authors, the positive effect of an attractive title on publication outcomes persists independently after accounting for author-specific factors. This finding underscores the importance of title choice in the publication process, even when controlling for author heterogeneity.

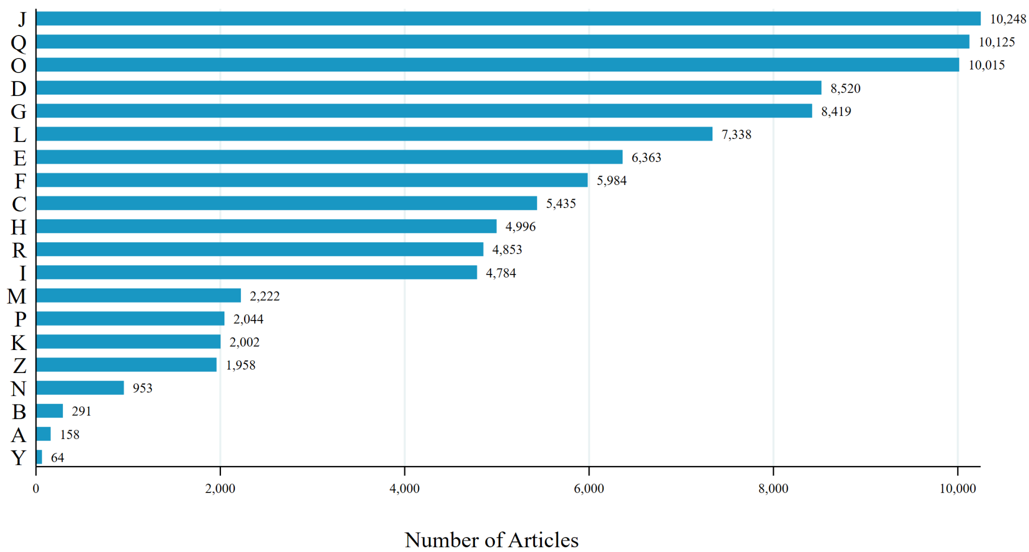
Columns (4) to (6) examine the impact of title attractiveness on citation counts. In Columns (4) and (5), we again alter the measure of author ability, but the regression results remain largely unaffected, and our findings continue to hold robustly. Interestingly, the coefficients for both the corresponding author's H-index and paper count are negative. This seemingly counterintuitive result may be explained by a diminishing marginal effect: for scholars with already high H-indices or numerous publications, new papers may face challenges in attracting citations at the same rate as their earlier works. This could be due to their previous contributions having already established foundational knowledge in their field, with newer works often representing incremental advancements.

In Column (6), after controlling for individual fixed effects of corresponding authors, the coefficient for title attractiveness becomes statistically insignificant. This change in significance likely stems from the introduction of multicollinearity due to the large number of control variables. We observe a substantial increase in the standard error of the title attractiveness variable, which primarily accounts for the loss of statistical significance. However, it's crucial to note that the direction and magnitude of the coefficient remain consistent with our previous findings. Furthermore, the citation process in academia typically prioritizes journal quality over individual author characteristics, especially after controlling for journal fixed effects. This suggests that author fixed effects may have limited additional explanatory power in determining citation patterns. Consequently, while the coefficient in Column (6) loses statistical significance, we contend that this does not fundamentally alter our baseline conclusions regarding the impact of title attractiveness on citations.

V.3 Including JEL

A potential concern with our analysis is that the trained model might exhibit systematic preferences for titles in specific research fields because of potential bias in the human annotator, potentially biasing our estimates. To address this concern, we introduce controls for research fields using JEL (Journal of Economic Literature) classification codes. Figure 6 presents the distribution of JEL codes in our sample.

FIGURE 7: Original JEL Disturbution



Our initial approach involves collecting JEL codes directly from the RePEc database for our sample articles. However, this direct collection method yields JEL codes for only 86,888 articles, representing approximately 26.7% of our total sample. As shown in Figure 6, the distribution of these JEL codes reveals substantial variation across economic subfields, with certain areas such as Labor and Demographic Economics (J), Agricultural and Natural Resource Economics • Environmental and Ecological

Economics (Q), and Economic Development, Innovation, Technological Change, and Growth (O) being more heavily represented.

To overcome this data limitation and maintain our full sample size, we employ a machine learning approach to predict JEL codes for the entire sample. Specifically, we utilize DeBERTa (Decoding-enhanced BERT with Disentangled Attention), a state-of-the-art transformer model that enhances BERT's architecture by disentangling attention mechanisms and introducing enhanced position encoding. This model has demonstrated superior performance in various natural language processing tasks compared to traditional BERT models [6].

We train the DeBERTa-large model using articles' titles and abstracts to predict their JEL codes. This approach aligns with Ash and Hansen (2023) framework for using BERT-type models to predict metadata in economics research. The intuition behind this approach is sound: titles and abstracts typically contain sufficient information to infer a paper's primary research fields. However, the task presents unique challenges due to its multi-label nature - papers can be assigned multiple JEL codes, and the number of assigned codes varies across papers. Moreover, some papers might be relevant to multiple fields but only report a subset of applicable JEL codes, potentially complicating the model's training process.

Despite these challenges, our DeBERTa model achieves strong predictive performance after training for 35,000 steps with a batch size of 4. The model attains an F1 score of 0.74 and a Jaccard index of 0.665 on the validation set, indicating that our predicted JEL codes overlap with actual codes by approximately two-thirds. This performance level suggests that the model captures a substantial portion of the true field classifications.

[6] You can get the raw model through this website: https://huggingface.co/microsoft/deberta-large

TABLE 6: Robustness Estimation (JEL)

	Orginal JEL		Predicted JEL	
	(1)	(2)	(3)	(4)
Eye Catching	0.106***	3.636**	0.090***	2.883***
	(0.009)	(1.541)	(0.006)	(0.746)
Corr. Citations	0.000***	0.001***	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.000***	0.006***	0.000***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.509***		0.534***	
	(0.011)		(0.005)	
General	0.457***		0.210***	
	(0.008)		(0.004)	
Year FE	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	86269	86888	297997	298983

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6 presents regression results using both the original JEL codes from RePEc in Columns (1) - (2) and the predicted JEL codes from our DeBERTa model in Columns (3) - (4). For journal quality, comparing Columns (1) and (3), we observe that the coefficient on title attractiveness remains positive and significant across both specifications. Similarly, for citations in Columns (2) and (4), the effect remains robust, though slightly attenuated when using predicted JEL codes.

The robustness of our results after controlling for JEL codes is particularly noteworthy. It demonstrates that even after accounting for potential field-specific biases in our human annotators' preferences, the positive effects of title attractiveness on both publication outcomes and citation counts persist. This suggests that our findings capture a fundamental relationship between title attractiveness and academic success that transcends individual research fields, rather than reflecting the idiosyncratic preferences of annotators in particular economic subfields.

VI. Further Discussion

VI.1 Hetergeniety of Journal Quality

Table 7: Hetergeneity: ABS ranking

	ABS:1	ABS:2	ABS:3	ABS:4	ABS:4*
	(1)	(2)	(3)	(4)	(5)
Eye Catching	-0.037	-0.675	3.710***	13.748**	3.137
	(0.390)	(0.505)	(1.033)	(5.735)	(9.600)
Corr. Citations	0.000***	0.001***	0.001***	0.002***	0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Team Citations	0.001***	0.002***	0.004***	0.006***	0.010***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	57295	116341	92955	26437	7967

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In this section, we further examine the impact of title attractiveness on citation counts across journals of varying quality, as classified by different journal ranking systems. Table 7 presents the results of our regressions using the ABS Journal Guide. From Columns (1) to (5), we observe significant heterogeneity in the effect of title attractiveness across different journal tiers. Overall, we find an inverted U-shaped relationship between title attractiveness and citation counts.

Specifically, for articles published in 1- or 2-star journals, title attractiveness has no statistically significant effect on citation counts. However, for articles in 3- and 4-

star journals, higher title attractiveness is associated with significantly higher citations, with the effect being more pronounced for 4-star journals. Interestingly, for the top-tier 4* journals, the effect of title attractiveness on citations diminishes and becomes statistically insignificant.

These findings align with our intuition. When articles are published in lower-quality journals, readers tend to focus on the quality of the paper itself rather than being swayed by an attractive title. If the paper is perceived as lacking in quality or relevance, readers are likely to disregard it, regardless of how compelling the title might be. In contrast, for articles published in higher-quality but not top-tier journals, readers may face a trade-off—they are more inclined to pay attention to titles that capture their interest since the overall quality of the journal guarantees a certain standard. Readers are less likely to thoroughly examine every paper, and more attractive titles help them decide which papers to explore further. However, for articles published in the very top journals (top 5), the title's attractiveness becomes less important. Scholars are often incentivized to cite top-tier papers simply because doing so is beneficial for their own work, as these papers are widely regarded as high quality. As long as there is some relevance between a top-tier paper and their own work, scholars are likely to cite it.

TABLE 8: Hetergeneity: Tinbergen Institute Ranking

	AA	A	B	Other
	(1)	(2)	(3)	(4)
Eye Catching	3.148	8.325***	2.861***	0.240
	(9.599)	(2.839)	(0.759)	(0.699)
Corr. Citations	0.000	0.001***	0.001***	0.001***
	(0.001)	(0.000)	(0.000)	(0.000)
Team Citations	0.010***	0.006***	0.003***	0.002***
	(0.001)	(0.001)	(0.000)	(0.000)
Year FE	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes
Affiliation Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	7965	34682	80990	177358

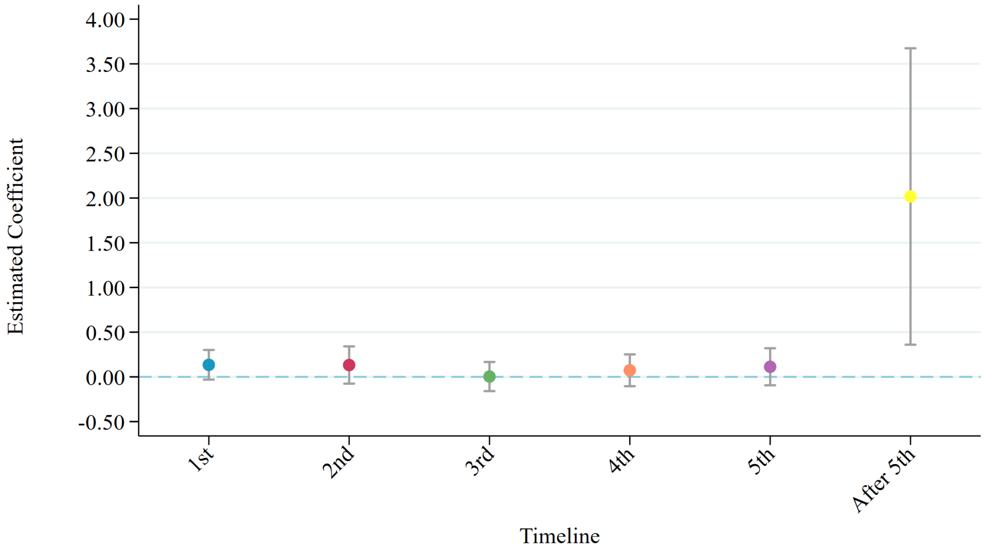
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also conducted a similar heterogeneity analysis using the Tinbergen Institute Rank, and the results are presented in Table 8. We find that these results are largely consistent with those obtained using the ABS Journal Guide. Specifically, the effect of title attractiveness on citations is mainly concentrated in journals ranked as A and B, which is entirely in line with our expectations.

VI.2 Hetergeniety of Citation Pattern on Different Years

FIGURE 8: Citation Pattern on Different Years



In the previous section, we found that papers with higher title attractiveness tend to attract more citations, particularly when published in high-quality journals. This raises another question: does the effect of title attractiveness only manifest when a paper already has a substantial number of citations? To explore this question, we obtained detailed citation data from the OpenAlex Database, which records yearly citation counts. We transformed this data into first-year citations, second-year citations, and so on, up to citations after the fifth year, treating each as a flow rather than a stock. Figure 7 presents the estimated impact of title attractiveness on citation rates over time.

Our estimation reveals that, in the first five years after publication, the estimated effects of title attractiveness are quite similar, with small coefficients around 0.25. Notably, the effect in the first year is slightly larger and statistically significant at the 10% level, while the effects in other years are not statistically significant. This suggests a modest initial advantage, indicating that the effect of more attractive titles can work even when the paper has not yet accumulated other citations. Papers with higher title attractiveness are more likely to attract citations in the early stages compared to other papers.

A distinct trend emerges in the period labeled as "After 5th year." In this later stage, the estimated coefficient is substantially larger, closer to 2, indicating a notable increase in citations for papers with more attractive titles over the long term. This reflects a cumulative effect, combined with the impact of title attractiveness. It implies that papers with higher title attractiveness not only attract citations early on but also accumulate enough citations to become even more noticeable (due to the cumulative citation effect), thereby enhancing the impact of the attractive title over time. This suggests that a slight advantage in title attractiveness at the beginning can lead to a significant gain in citations in the future.

VI.3 Other controls

TABLE 9: Robustness Estimation (Other Controls)

	Novelty		Title Length	
	(1)	(2)	(3)	(4)
Eye Catching	0.134***	1.409*	0.128***	1.911***
	(0.006)	(0.765)	(0.005)	(0.734)
Corr. Citations	0.000	0.001***	-0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.000***	0.004***	0.000***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.552***		0.533***	
	(0.004)		(0.004)	
General	0.174***		0.181***	
	(0.005)		(0.004)	
Novelty Index	0.176***	2.271		
	(0.011)	(1.549)		
Title Length			-0.022***	-0.156***
			(0.000)	(0.057)
Year FE	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	271627	272492	300006	300995

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To further refine our analysis and address potential confounding factors, we introduce two additional controls: article novelty and title length. Table 9 presents the results of these extended specifications.

We measure article novelty using the method developed by Bramoullé and Ductor (2018), which quantifies the atypicality of keyword combinations in a paper relative to previously published articles. This approach captures the innovative aspects of research by assessing the uniqueness of its central themes. Specifically, the novelty index is calculated as the normalized average atypicality of keyword pairs, where atypicality is measured by the negative log probability of observing a given keyword pair in prior literature. This index ranges from 0 to 1, with values closer to 1 indicating higher novelty.

Columns (1) and (2) of Table 9 show the results when controlling for novelty. In Column (1), we observe that the coefficient on "Eye Catching" remains positive and significant at 0.134, while the novelty index itself shows a positive and significant effect. This suggests that both title attractiveness and article novelty independently contribute to publication in higher-quality journals. In Column (2), we see a similar pattern for citations, with title attractiveness maintaining a positive and significant effect, although the novelty index is not statistically significant in this specification. This indicates that novelty may not be a primary consideration in the citation process. These results are particularly interesting as they suggest that while novel research is more likely to be published in higher-quality journals, it does not necessarily attract more citations. This could reflect a tension between cutting-edge research that pushes boundaries (appealing to top journals) and work that builds on established literature (potentially garnering more citations).

We also control for title length, motivated by Bramoullé and Ductor (2018) finding that title length can influence both publication outcomes and citation counts. Given the potential correlation between eye-catching titles and brevity, we include this control to ensure the robustness of our results. Columns (3) and (4) present these findings.

In Column (3), we find that title attractiveness remains positive and significant , while title length has a negative and significant effect. Similarly, In Column (4), title attractiveness maintains a positive and significant effect, with title length showing a negative and significant impact. These results align with Bramoullé and Ductor (2018) findings, suggesting that shorter titles are associated with both publication in higher-quality journals and increased citation counts. The persistence of the positive and significant effect of title attractiveness, even after controlling for title length, is particularly noteworthy. It indicates that the impact of an eye-catching title goes beyond mere brevity, encompassing qualitative aspects that independently influence a paper's success. This finding underscores the multifaceted nature of effective academic communication.

The robustness of our title attractiveness measure to these additional controls provides further evidence of its importance in academic publishing. It suggests that crafting an attractive title is a distinct skill from writing concisely or conducting novel research, and that all these factors play independent roles in determining a paper's success.

VI.4 Using Bert Logits

TABLE 10: Robustness: Using Logits and Separate Regression

	Journal Quality					Citation				
	Logits	Thres.:0.6	Thres.:0.7	Thres.:0.8	Thres.:0.9	Logits	Thres.:0.6	Thres.:0.7	Thres.:0.8	Thres.:0.9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Eye Catching Logits	0.152***					2.283***				
	(0.006)					(0.818)				
Eye Catching		0.130***	0.131***	0.131***	0.134***		2.110***	2.117***	2.232***	2.754***
		(0.006)	(0.006)	(0.006)	(0.006)		(0.751)	(0.773)	(0.786)	(0.857)
Correspond Citations	-0.000	-0.000	-0.000	-0.000	-0.000	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.000***	0.000***	0.000***	0.000***	0.000***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.564***	0.562***	0.561***	0.561***	0.560***					
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)					
General	0.193***	0.193***	0.193***	0.193***	0.193***					
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)					
Year FE	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Journal FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	300006	300006	300006	300006	300006	300995	300995	300995	300995	300995

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In our primary analysis, we used a binary classification of title attractiveness derived from our fine-tuned BERT model. However, to provide a more nuanced understanding and test the robustness of our findings, we employ alternative measures based on the raw output of our BERT model: the logits.

Before delving into the analysis, it's important to understand what these logits represent. In a binary classification task, such as ours, the BERT model processes the title through multiple layers and ultimately produces a single number, the logit. This logit is then passed through a sigmoid function to produce a probability between 0 and 1, representing the likelihood of the title being classified as "eye-catching". In our main analysis, we used a threshold of 0.5 on this probability to create our binary classification: titles with probabilities above 0.5 were classified as eye-catching (1), while those below were not (0).

Using these logits, we conduct two types of regressions. First, we use the raw logits as a continuous measure of title attractiveness. Second, we create multiple binary variables using different probability thresholds (0.6, 0.7, 0.8, and 0.9) to examine how the effect of title attractiveness varies at different levels of ”eye-catchingness”.

Table 10 presents the results of these analyses. In Columns (1) and (6), we replace our binary ”Eye Catching” variable with the continuous ”Eye Catching Logits” measure. The results are consistent with our main findings, showing a positive and statistically significant relationship between title attractiveness and both journal quality and citation counts. Specifically, a one-unit increase in the logit score is associated with a 0.152 increase in journal quality (measured by ABS star rating) and 2.283 additional citations, both significant at the 1% level.

Columns (2)-(5) and Columns (7)-(10) present the results of our threshold analysis. For journal quality (Columns 2-5), we observe an interesting pattern: the coefficient on our eye-catching measure increases as we raise the threshold, from 0.130 at the 0.6 threshold to 0.134 at the 0.9 threshold. This suggests that titles classified as eye-catching at higher thresholds are associated with publication in higher-quality journals. However, the differences between these coefficients are relatively small, indicating a fairly consistent effect across different levels of title attractiveness.

For citations (Columns 7-10), we see a more pronounced pattern. The coefficient increases from 2.110 at the 0.6 threshold to 2.754 at the 0.9 threshold. This substantial increase suggests that titles classified as highly attractive (i.e., those clearing the 0.9 threshold) are associated with significantly more citations than those just barely classified as eye-catching.

These findings provide several important insights. First, they confirm the robustness of our main results: regardless of how we measure title attractiveness, we consistently find a positive relationship with both publication quality and citation counts. Second, they suggest a potential non-linear relationship, particularly for citations, where the most attractive titles seem to yield disproportionately large benefits.

The increasing coefficients as we raise the threshold are particularly intriguing, especially when we compare the effects on journal quality and citations. For journal quality, the relatively small increases suggest that while more attractive titles are associated with publication in better journals, this effect is fairly consistent across different levels of attractiveness. For citations, however, the larger increases in coefficients suggest that there might be substantial returns to crafting particularly eye-catching titles.

This difference in the marginal effects of title attractiveness on publication outcomes and citation counts is noteworthy and warrants further discussion. In the publication process, title attractiveness likely serves as an initial impression, but it is subsequently complemented by the paper's content, methodology, and results during the rigorous peer review process. Once a title crosses a certain threshold of attractiveness, its marginal benefit in the review process may diminish as other factors take precedence. This could explain the relatively consistent effect we observe across different levels of title attractiveness for journal quality. In contrast, the citation process operates differently. Researchers often rely on titles, abstracts, and introductions when deciding which papers to cite, particularly when dealing with a large volume of potentially relevant literature. In this context, an exceptionally eye-catching title may carry more weight, as it can significantly influence whether a paper is read and subsequently cited. This could explain the more pronounced increase in the effect of title attractiveness on citations as we move to higher thresholds.

VI.5 Professor Preference and Student Preference

TABLE 11: Robustness: Prof vs Student

	Professor				Student			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eye Catching	0.067***	2.226***			0.008*	1.410**		
	(0.005)	(0.672)			(0.005)	(0.553)		
Eye Catching Logits			0.133***	4.068***			0.011**	1.506**
			(0.008)	(1.118)			(0.005)	(0.647)
Correspond Citations	-0.000	0.001***	-0.000	0.001***	-0.000	0.001***	-0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.000***	0.004***	0.000***	0.004***	0.000***	0.004***	0.000***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.558***		0.565***		0.552***		0.553***	
	(0.004)		(0.004)		(0.004)		(0.004)	
General	0.194***		0.194***		0.194***		0.194***	
	(0.004)		(0.004)		(0.004)		(0.004)	
hline Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes	No	Yes	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	300006	300995	300006	300995	300006	300995	300006	300995

Standard errors in parentheses

sym* (p¡0.05), sym** (p¡0.01), sym*** (p¡0.001)

The perception of title attractiveness may vary between established scholars and early-career researchers, potentially affecting our understanding of its impact on academic outcomes. To investigate this possibility and provide an additional robustness check for our main results, we conducted separate analyses using BERT models trained on professors' and students' evaluations of title attractiveness. Table 11 presents the results of this analysis. Columns (1) - (4) show the outcomes for the professor-trained model, while Columns (5) - (8) display results for the student-trained model. For each group, we examine the impact on both journal quality and citation counts, using binary classification and continuous logit scores as measures of title attractiveness.

Examining journal quality first, we find that both professor- and student-trained models yield positive and statistically significant coefficients for title attractiveness. However, the magnitude of the effect is notably larger for the professor-trained model. In Column (1), the professor model shows a 0.067 increase in journal quality for attractive titles, compared to a 0.008 increase for the student model in Column (5). This pattern persists when using continuous logit scores, with coefficients of 0.133 and 0.011 for professor and student models, respectively, as shown in Columns (3) and (7).

For citation counts, we again observe positive and significant effects across all specifications, but with more pronounced differences between professor and student models. In Column (2), the professor-trained model indicates that papers with attractive titles receive 2.226 more citations on average, while the student-trained model suggests an increase of 1.410 citations in Column (6). These differences are even more striking when using continuous logit scores, with coefficients of 4.068 and 1.506 for professor and student models, respectively, as shown in Columns (4) and (8).

These findings have several important implications for the robustness of our results and our understanding of title attractiveness. First, the consistency in the direction and significance of the effects across both professor and student models reinforces the robustness of our main findings. This is particularly important given the potential concern that professors, due to their long-term exposure to economic literature, might inadvertently incorporate characteristics of top-tier journal titles into their assessment of "eye-catching" titles. While we cannot completely eliminate this issue, as these features are inherently mixed in professors' recognition patterns, the use of a student-trained model allows us to partially address this concern. The fact that the student-trained model also yields positive and significant results provides strong evidence for the robustness of our findings.

Second, the larger coefficients observed in the professor-trained model suggest that experienced scholars perceive a stronger predictive power of attractive titles for both publication success and citation counts. This discrepancy could be attributed to two factors. On one hand, it may reflect the possibility that professors' recognition of eye-catching titles is indeed intertwined with their ability to identify characteristics of top-tier journal publications. On the other hand, it could also indicate that professors, as the primary reviewers and citing authors in academia, have a more accurate understanding of which title characteristics are likely to lead to publication success and academic impact.

VI.6 Using Generative AI in Title Evaluation

TABLE 12: Robustness: Other preference measured by LLMs (Normal Prompts)

	CHATGPT 3.5		LLAMA3-8B		GPT-4O	
	(1)	(2)	(3)	(4)	(5)	(6)
Eye Catching	0.066***	5.947***	0.071***	1.664	0.121***	3.696**
	(0.012)	(1.323)	(0.011)	(1.314)	(0.015)	(1.850)
Corr. Citations	-0.000	0.001***	-0.000	0.001***	-0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.000***	0.003***	0.000***	0.003***	0.000***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.568***		0.567***		0.565***	
	(0.013)		(0.013)		(0.013)	
General	0.197***		0.197***		0.195***	
	(0.014)		(0.014)		(0.014)	
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30149	30247	30149	30247	30149	30247

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To further validate our findings and explore innovative methodologies in economics research, we employ three generative artificial intelligence (GenAI) models to assess title attractiveness: ChatGPT 3.5, LLAMA3-8B, and GPT-4. These models represent a spectrum of AI capabilities: ChatGPT 3.5 serves as a baseline model, LLAMA3-8B represents a high-performing open-source model, and GPT-4 exemplifies a state-of-the-art model. This approach serves several purposes. First, it provides a robust check on our main results by leveraging a different type of aggregated preference. GenAI models, trained on vast datasets generated by diverse individuals, represent a distinct composite preference that can serve as a valuable counterpoint to our human-annotated sample, providing a supplementary check to mitigate potential biases in our human-labeled sample. Second, it explores the feasibility of using GenAI as a substitute for human subjects in economics experiments which require humans to have different preferences. Our study employed a complex and time-consuming process of human annotation and spent considerable time fine-tuning a BERT model. However, if GenAI models can closely approximate human composite preferences, it could significantly streamline future research methodologies in this area. Third, it investigates the potential of GenAI in evaluating title attractiveness for practical applications. If GenAI judgments align closely with those of human experts, researchers could potentially use these models to assess the attractiveness of their own paper titles, thereby potentially improving publication success rates and citation counts.

Due to budget constraints, we conducted this analysis on a stratified random sample of 10% of our total dataset. While this limitation reduces statistical power, it allows us to maintain cost-effectiveness while still providing valuable insights.

Tables 12 and 13 present the results of our analysis using GenAI models to evaluate title attractiveness. In Table 12, we use a prompt similar to that given to our human annotators, while Table 13 employs a prompt that instructs the AI to assume the perspective of an experienced economist who read a lot of paper.

Examining Table 12, we observe that the coefficients for title attractiveness remain positive and statistically significant across all specifications for ChatGPT 3.5 and GPT-4. For journal quality (columns 1 and 5), the coefficients are 0.066 and 0.121 respectively. For citations (columns 2 and 6), we see positive and significant effects, with coefficients of 5.947 and 3.696. Interestingly, while LLAMA3-8B shows a significant positive effect on journal quality (0.071), its effect on citations is not statistically significant. This reflects that the aggregate preference presented by LLAMA3-8B may not be as effective at identifying eye-catching titles that attract readers' attention.

We also observe a trade-off between the publication prospects and citation rates of eye-catching titles. While the titles GPT-4 considers eye-catching have higher publication prospects, they have relatively lower citation rates. This may reflect differences in preferences between editors and reviewers versus those who cite the papers. Editors and reviewers are typically experts in the field, carefully selected and more attuned to field-specific preferences and standards. In contrast, those citing papers may come from diverse fields and face fewer barriers to citation. This could lead to a trend where titles meeting the "eye-catching" standards of seasoned scholars in a specific field may not align as well with general interest, resulting in the observed trade-off between publication prospects and citations. This phenomenon might also explain why some papers published in top 5 journals don't receive as many citations as some papers in some normal journals.

TABLE 13: Robustness: Other preference measured by LLMs (Economists Prompts)

	CHATGPT 3.5		LLAMA3-8B		GPT-4O	
	(1)	(2)	(3)	(4)	(5)	(6)
Eye Catching	0.044***	7.036***	0.098***	2.706*	0.109***	4.865***
	(0.011)	(1.479)	(0.011)	(1.402)	(0.011)	(1.324)
Corr. Citations	-0.000	0.001***	-0.000	0.001***	-0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Team Citations	0.000***	0.003***	0.000***	0.003***	0.000***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Theoritical	0.568***		0.581***		0.579***	
	(0.014)		(0.014)		(0.014)	
General	0.197***		0.196***		0.196***	
	(0.014)		(0.014)		(0.014)	
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30149	30247	30149	30247	30149	30247

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Given these observations, we wondered whether prompting the model to assume the perspective of an economist who has read many papers might shift the model's preferences towards those of high-level economics editors and reviewers. Table 13, which uses the economist-framed prompt, shows similar patterns but with some notable differences. For ChatGPT 3.5, the economist prompt leads to a smaller coefficient for journal quality (0.044 vs 0.066) but a larger coefficient for citations (7.036 vs 5.947).

LLAMA3-8B shows a larger coefficient for journal quality with the economist prompt (0.098 vs 0.071) and, notably, a significant effect on citations (2.706) which was not present with the neutral prompt. GPT-4 demonstrates more consistent results across prompts, with slightly lower coefficients for the economist prompt.

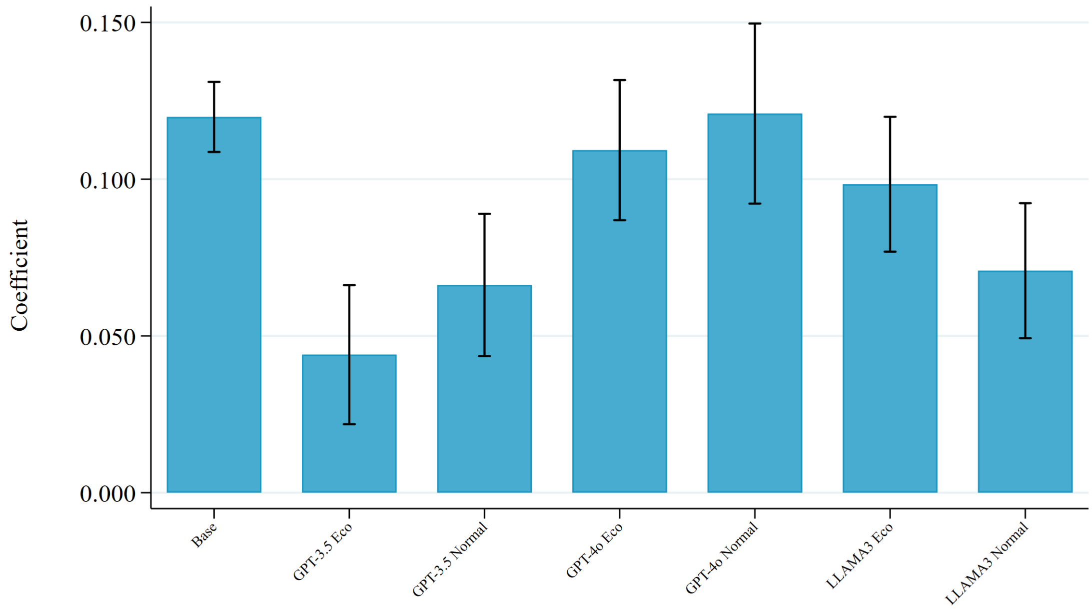
Contrary to our expectations, the economist prompt did not consistently adjust the models' preferences towards those of high-level editors and reviewers. While LLAMA3-8B showed improvements in both journal quality and citation coefficients and significance, the other models, particularly for citations, showed increased coefficients. This suggests that the prompt may have emphasized the model's role as a reader rather than an editor or reviewer. These results highlight the challenges and potential unpredictability of using prompt engineering to fine-tune model preferences, indicating that careful consideration is necessary when attempting to align AI models with specific human preferences.

These results provide strong support for the robustness of our main findings. The consistency of positive and significant effects across these different GenAI models suggests that our results are not artifacts of specific human biases or limited sample sizes. Moreover, the statistical significance across most models indicates the potential feasibility of using GenAI as a substitute for human subjects in preference experiments of this nature. However, the results also caution against the uncontrolled use of prompt engineering to approximate specific types of human preferences, highlighting the need for careful and rigorous testing in such applications.

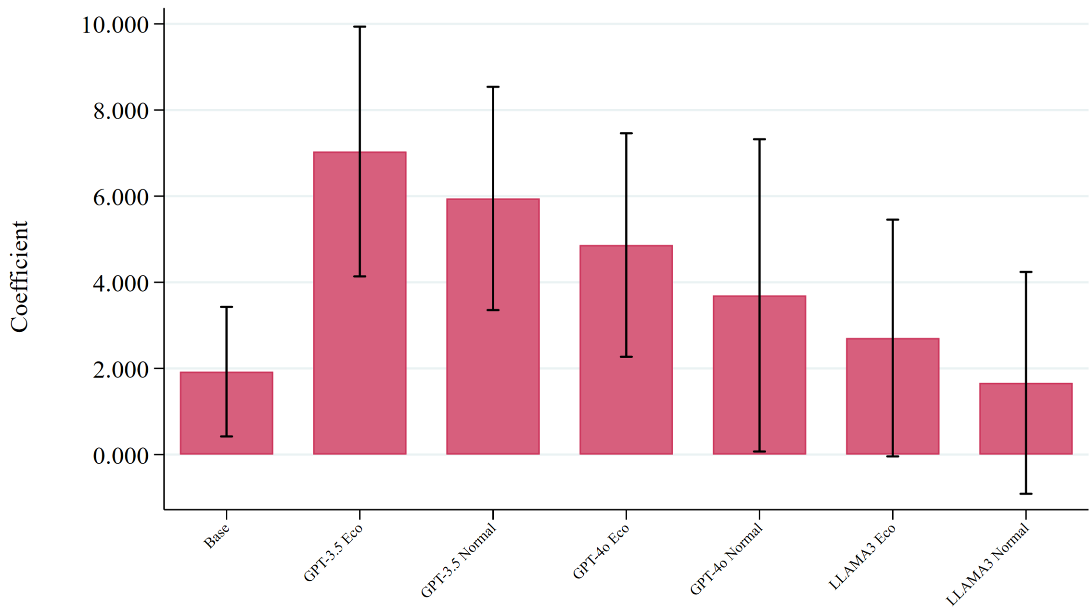
These differences also highlight the nuanced nature of title attractiveness and suggest that different aspects of attractiveness may be more or less important depending on the outcome of interest (journal quality vs. citations) and the specific AI model used. This insight could be valuable for researchers in crafting titles that are optimized for their specific goals and in selecting appropriate AI tools for analysis.

FIGURE 9: Coefficient Comparison on Different Models

(A) ABS Journal Guide



(B) Citations



To provide a more comprehensive comparison, Figure 8 presents a visual representation of the coefficients from our main human-annotated model alongside those from the various GenAI models.

Panel A of Figure 8 compares the coefficients for journal quality across different models. We find that GPT-4O's estimated coefficients are closest to those of the BERT model fine-tuned on human samples, suggesting that more capable models have preference recognition abilities in specialized fields that approach those of human experts. Conversely, ChatGPT 3.5 performs the poorest, implying that its preferences are closer to those of the general public, aligning more with general interests in its judgments. This logic is consistent with its larger coefficient for citations. LLAMA3 falls between these two models in terms of performance.

Panel B presents a similar comparison for citation counts. Here, we observe a different pattern. ChatGPT 3.5, particularly with the economist prompt, shows the highest coefficient, even surpassing our human-annotated model. This suggests that ChatGPT 3.5 may be particularly adept at identifying title characteristics that correlate with higher citation rates. GPT-4O, while performing well, shows a more conservative estimate compared to the human-annotated model. LLAMA3 again falls between the two, but with a notably lower coefficient compared to its performance on journal quality.

These observations highlight an interesting dichotomy: while more advanced models like GPT-4O seem to better align with expert judgments on publication quality, simpler models like ChatGPT 3.5 may be more attuned to features that drive citations. This could reflect the different nature of these two outcomes - publication quality being more tied to expert assessment, while citations may be influenced by broader appeal factors that ChatGPT 3.5 captures well.

The proximity of coefficients between GenAI models and our human-annotated model, particularly for citation counts, further supports the feasibility of using GenAI as a substitute for human subjects in preference experiments. It also helps us identify which models and prompts might be most suitable for this type of research. For instance, the strong performance of ChatGPT 3.5 with the economist prompt in predicting citation impacts suggests it might be particularly useful for researchers aiming to optimize their titles for citation potential.

Moreover, these results underscore the importance of model selection in such analyses. Depending on the specific research question or outcome of interest, different models may be more or less appropriate. For studies focused on predicting publication in high-quality journals, more advanced models like GPT-4O might be preferred. In contrast, for research aimed at maximizing citation potential, ChatGPT 3.5 with appropriate prompting could be the tool of choice.

These findings also have implications for the broader field of AI-assisted research in economics and other disciplines. They suggest that while AI models can provide valuable insights and potentially streamline certain research processes, their use requires careful consideration and validation against human expert judgments. The variation in performance across models and tasks highlights the need for researchers to thoroughly test and validate AI tools in their specific contexts before relying on them for important decisions or analyses.

VII. Conclusion

This paper investigates how title attractiveness affects publication outcomes and citation impacts in economics research. Our analysis yields several key findings that advance our understanding of academic publishing dynamics and research dissemination patterns.

First, we establish a robust causal relationship between title attractiveness and academic success. Papers with attractive titles are published in higher-ranked journals and receive more citations, even after controlling for comprehensive sets of author, institutional, and journal characteristics. The persistence of these effects across various specifications and robustness checks suggests that title crafting plays a meaningful role in determining research impact.

Second, we document significant heterogeneity in the impact of title attractiveness across journal tiers. The effect is particularly pronounced for mid-tier journals, while being less significant for both lower-ranked and top-tier outlets. This non-linear pattern suggests that title attractiveness serves as a valuable attention-capturing device in the increasingly competitive middle segment of academic publishing, where papers must actively compete for reader attention.

Third, our analysis of citation patterns reveals that the impact of title attractiveness extends beyond initial publication success. The effect on citations grows stronger over time, suggesting that attractive titles contribute to the cumulative advantage process in academic impact. This finding has important implications for understanding how initial attention advantages translate into long-term scholarly influence.

Fourth, our innovative application of machine learning techniques to assess subjective paper characteristics opens new methodological possibilities for empirical research in economics. The strong performance of both our fine-tuned BERT model and various large language models suggests that AI tools can effectively capture and quantify subjective academic judgments at scale.

These findings have several important implications. For individual researchers, our results suggest that investing time in crafting attractive titles can yield meaningful returns in terms of publication success and scholarly impact. For journal editors and reviewers, our findings highlight the need to consider whether title attractiveness might inadvertently influence evaluation processes. For the broader academic community, our work underscores the evolving nature of academic communication in an era of increasing competition for attention.

Our study also points to several promising directions for future research. First, investigating whether similar patterns exist in other disciplines could illuminate field-specific differences in how title characteristics influence academic success. Second, exploring how title attractiveness interacts with other paper characteristics, such as abstract quality or methodological sophistication, could provide a more complete picture of research impact determinants. Finally, examining whether the growing use of AI tools in title generation affects these dynamics represents an important area for future investigation.

REFERENCES

ALEXOPOULOS, M., K. LYONS, K. MAHETAJI, M. E. BARNES, AND R. GUTWILLINGER (2023): "Gender inference: can chatGPT outperform common commercial tools?" in *Proceedings of the 33rd Annual International Conference on Computer Science and Software Engineering*, 161–166.

ASH, E. AND S. HANSEN (2023): "Text algorithms in economics," *Annual Review of Economics*, 15, 659–688.

BRAMOULLÉ, Y. AND L. DUCTOR (2018): "length," *Journal of Economic Behavior & Organization*, 150, 311–324.

BRANSCH, F. AND M. KVASNICKA (2022): "Male gatekeepers: gender bias in the publishing process?" *Journal of Economic Behavior & Organization*, 202, 714–732.

CARD, D. AND S. DELLAVIGNA (2013): "Nine facts about top journals in economics," *Journal of Economic literature*, 51, 144–161.

CHAN, H. F., A. S. ÖNDER, S. SCHWEITZER, AND B. TORGLER (2023): "Twitter and citations," *Economics letters*, 231, 111270.

CHEN, Y., T. X. LIU, Y. SHAN, AND S. ZHONG (2023): "The emergence of economic rationality of GPT," *Proceedings of the National Academy of Sciences*, 120, e2316205120.

DELL, M. (2024): "Deep Learning for Economists," Working Paper 32768, National Bureau of Economic Research.

DEVLIN, J., M.-W. CHANG, K. LEE, AND K. TOUTANOVA (2018): "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*.

DUCTOR, L. AND B. VISSER (2022): "When a coauthor joins an editorial board," *Journal of economic behavior & organization*, 200, 576–595.

FELD, J., C. LINES, AND L. ROSS (2024): "Writing matters," *Journal of Economic Behavior & Organization*, 217, 378–397.

GALIANI, S., R. H. GÁLVEZ, AND I. NACHMAN (2023): "Unveiling specialization trends in economics research: A large-scale study using natural language processing and citation Analysis," Tech. rep., National Bureau of Economic Research.

GNEWUCH, M. AND K. WOHLRABE (2017): "Title characteristics and citations in economics," *Scientometrics*, 110, 1573–1578.

GORODNICHENKO, Y., T. PHAM, AND O. TALAVERA (2023): "The voice of monetary policy," *American Economic Review*, 113, 548–584.

GUO, F., C. MA, Q. SHI, AND Q. ZONG (2018): "Succinct effect or informative effect: The relationship between title length and the number of citations," *Scientometrics*, 116, 1531–1539.

HADAVAND, A., D. S. HAMERMESH, AND W. W. WILSON (2024): "Publishing economics: How slow? Why slow? Is slow productive? How to fix slow?" *Journal of Economic Literature*, 62, 269–293.

HAMERMESH, D. S. (2018): "Citations in economics: Measurement, uses, and impacts," *Journal of Economic Literature*, 56, 115–156.

HANSEN, S., P. J. LAMBERT, N. BLOOM, S. J. DAVIS, R. SADUN, AND B. TASKA (2023): "Remote work across jobs, companies, and space," Tech. rep., National Bureau of Economic Research.

HASAN, S. A. AND R. V. BREUNIG (2021): "Article length and citation outcomes," *Scientometrics*, 126, 7583 – 7608.

HECKMAN, J. J. AND S. MOKTAN (2020): "Publishing and promotion in economics: The tyranny of the top five," *Journal of Economic Literature*, 58, 419–470.

HORTON, J. J. (2023): "Large language models as simulated economic agents: What can we learn from homo silicus?" Tech. rep., National Bureau of Economic Research.

HOTBLLINO, H. (1929): "Stability in competition," *The economic journal*, 39, 41–57.

JELVEH, Z., B. KOGUT, AND S. NAIDU (2024): "Political language in economics," *The Economic Journal*, ueae026.

KINNEY, R., C. ANASTASIADES, R. AUTHUR, I. BELTAGY, J. BRAGG, A. BURACZYNSKI, I. CACHOLA, S. CANDRA, Y. CHANDRASEKHAR, A. COHAN, ET AL. (2023): "The semantic scholar open data platform," *arXiv preprint arXiv:2301.10140*.

KOVÁCS, B., G. HSU, AND A. SHARKEY (2024): "The Stickiness of Category Labels: Audience Perception and Evaluation of Producer Repositioning in Creative Markets," *Management Science*, 70, 6315–6335.

LEE, J., W. YOON, S. KIM, D. KIM, S. KIM, C. H. SO, AND J. KANG (2020): "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, 36, 1234–1240.

LOUGHRAN, T. AND B. MCDONALD (2011): "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of finance*, 66, 35–65.

MA, C., Y. LI, F. GUO, AND K. SI (2019): "The citation trap: Papers published at year-end receive systematically fewer citations," *Journal of Economic Behavior & Organization*, 166, 667–687.

MCCANNON, B. C. (2019): "Readability and research impact," *Economics Letters*, 180, 76–79.

PETTY, R. E. AND P. BRIÑOL (2011): "The elaboration likelihood model," *Handbook of theories of social psychology*, 1, 224–245.

PRIEM, J., H. PIWOWAR, AND R. ORR (2022): "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," *arXiv preprint arXiv:2205.01833*.

SHAPIRO, A. H., M. SUDHOF, AND D. J. WILSON (2022): "Measuring news sentiment," *Journal of econometrics*, 228, 221–243.

SIMS, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

SUN, C., X. QIU, Y. XU, AND X. HUANG (2019): "How to fine-tune bert for text classification?" in *Chinese computational linguistics: 18th China national conference, CCL 2019, Kunming, China, October 18–20, 2019, proceedings 18*, Springer, 194–206.

TVERSKY, A. AND D. KAHNEMAN (1973): "Availability: A heuristic for judging frequency and probability," *Cognitive psychology*, 5, 207–232.

——— (1981): "The framing of decisions and the psychology of choice," *science*, 211, 453–458.

WALKER, J. T., E. FENTON, A. SALTER, AND R. SALANDRA (2019): "What influences business academics' use of the association of business schools (ABS) list? Evidence from a survey of UK academics," *British Journal of Management*, 30, 730–747.

ZHANG, Z., K. YANG, J. Z. ZHANG, AND R. W. PALMATIER (2023): "Uncovering synergy and dysergy in consumer reviews: A machine learning approach," *Management Science*, 69, 2339–2360.