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

Aljoscha Janssen* Shen Zekai†

Tuesday 31st December, 2024

Abstract

We study the relationship between title attractiveness and 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 between 2000 and 2022. Papers with titles that are considered attractive are published in journals with ABS Academic Journal Guide rankings 0.120 points higher and receive 1.925 (7.5%) 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. Our findings provide evidence of how title characteristics influence academic success and demonstrate the potential of machine learning in analyzing subjective paper features at scale.

Title Attractiveness, Publication success, Citation, BERT, LLMs

JEL Codes: A11, A14, O30

*Singapore Management University, ajanssen@smu.edu.sg

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

I. INTRODUCTION

Academic publishing in economics has become increasingly competitive, with top journals' acceptance rates falling below 7% and publication times often exceeding two years (Card and DellaVigna, 2013; Hadavand et al., 2024). In this environment, publication success significantly impacts career prospects and scholarly reputation (Hamermesh, 2018; Heckman and Moktan, 2020).

As standards for contributions 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 presentation-related factors significantly influence publication success and citation impact. These factors span from article length (Hasan and Breunig, 2021) and writing style (Feld et al., 2024) to coauthor characteristics (Bransch and Kvasnicka, 2022; Ductor and Visser, 2022). Strategic decisions also matter, including submission timing (Ma et al., 2019) and post-publication dissemination (Chan et al., 2023). Even seemingly peripheral elements like abstract readability affect paper reception (McCannon, 2019), highlighting how various factors beyond core research quality shape academic success.

Among these factors, paper titles warrant particular attention as the initial point of contact between research and potential readers. Despite requiring minimal time investment, titles may substantially influence both publication outcomes and subsequent citations in economics (Bramoullé and Ductor, 2018; Gnewuch and Wohlrabe, 2017; Guo et al., 2018). Prior research has focused primarily on observable title characteristics—such as length and non- alphabetical characters—finding that shorter titles (Bramoullé and Ductor, 2018) and those with non-alphabetical characters (Gnewuch and Wohlrabe, 2017) correlate with improved publication success and higher citation rates.

While prior research links observable title characteristics to improved publication metrics (Bramoullé and Ductor, 2018; Gnewuch and Wohlrabe, 2017), it does not focus on a framework explaining how such features shape editorial decisions, reader engagement, and scholarly impact. This gap persists largely because measurable title traits do not directly capture subjective perceptions, and gathering these responses at scale is challenging.

We suggest viewing title attractiveness as a concept that, despite its subjective aspects, can be explained by the underlying language and style choices that shape readers' responses. This approach recognizes that while individual responses to titles may vary, these responses are shaped by observable textual elements – such as word choice, phrase structure, and rhetorical devices – that collectively affect how a title captures attention and frames research contributions.

Titles do not only affect readers but also editors and reviewers. Editors face a flood of submissions and must allocate limited attention efficiently (Sims, 2003). In this context, an attractive title acts as an initial filter, encouraging deeper engagement and sustained focus, even during lengthy review cycles. Likewise, reviewers encountering engaging titles may form more positive first impressions, potentially influencing subsequent evaluations (Tversky and Kahneman, 1981).

To answer the research question if title attractiveness increases publication success and subsequent citation counts, 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 relatively objective assessment of title attractiveness. We develop a novel two-stage methodology combining human evaluation with advanced natural language processing techniques. 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 Encoder Representations from Transformers) (Devlin et al., 2018) to learn from and synthesize the human annotations.[1]

The empirical analysis reveals systematic patterns in how title attractiveness influences academic outcomes. In the publication process, papers with attractive titles are published in journals with ABS ratings 0.120 points higher than those without such titles – an effect that reflects how title attractiveness can influence editorial decision-making under attention constraints. In terms of citations, attractive titles are associated with 1.925 additional citations per paper (representing a 7.5% increase). While this effect might appear modest in absolute terms, its persistence over time and across different journal tiers suggests that title attractiveness systematically influences how research disseminates through academic networks. Particularly in middle-tier journals, where papers compete intensely for scholarly attention, attractive titles can help valuable research reach broader audiences, with effects accumulating over papers' citation lifecycles. These patterns are maintained after controlling for an extensive set of author, institutional, and journal characteristics, and journal, institution, and author fixed effects. Moreover, the conclusions remain stable across alternative journal ranking systems.

Further investigation reveals that the impact of title attractiveness exhibits notable heterogeneity across journal tiers. While the effect is minimal for papers in lower-tier journals (one or two stars), it becomes substantial for those in middle and upper-middle-tier outlets (three stars or A-level), peaking at 13.7 additional citations for papers in 4-star journals. Interestingly, this effect moderates for publications in the most prestigious journals (4* or AA-level). This pattern suggests that title attractiveness matters most for papers in middle-tier journals.

The analysis of large language models (LLMs) in assessing title attractiveness yields promising results. Despite their distinct architectures and training data, evaluations from GPT-4, ChatGPT-3.5, and LLAMA3 demonstrate remarkable consistency with our BERT model's predictions in both direction and statistical significance, suggesting the potential of generative AI for subjective evaluations in economic research. Different LLMs exhibit distinct strengths that reflect their underlying capabilities: GPT-4's evaluations better predict publication outcomes in prestigious journals, while ChatGPT-3.5 demonstrates superior performance in predicting citation impact. This divergence appears to reflect the models' varying levels of specialization—GPT-4's preferences align more closely with expert reviewers in specialized

[1] BERT has already demonstrated its utility in various cutting-edge economics and management science research (Ash and Hansen, 2023), including sentiment analysis (Gorodnichenko et al., 2023), text classification (Zhang et al., 2023), and information extraction from unstructured data (Hansen et al., 2023).

fields, while ChatGPT-3.5's broader training better captures the preferences of the general academic readership.

This study makes three contributions to the literature on academic publishing in economics and scientometrics. First, we provide robust empirical evidence on how title attractiveness shapes publication outcomes and citation counts in economics. While previous research has examined correlations between observable title characteristics and academic success (Bramoullé and Ductor, 2018; Gnewuch and Wohlrabe, 2017; Guo et al., 2018), we demonstrate that title attractiveness, as a subjective measure, influences both publication venue prestige and subsequent citation rates beyond these observable features. Our analysis, grounded in theories of limited attention (Sims, 2003) and framing effects (Tversky and Kahneman, 1981), elucidates how nuanced aspects of titles affect publication and citation patterns, advancing our understanding of academic impact determinants 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 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, this study advances the emerging literature on the application of large language models (LLMs) in social science research (Chen et al., 2023; Horton, 2023). We demonstrate that assessments of title attractiveness derived from off-the-shelf LLMs yield results consistently comparable 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 these models can generate data of comparable quality to human judgments even for subjective tasks in non-experimental 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 Title Attractiveness

This section summarizes our data construction and the approach to measuring title attractiveness in economics publications. We describe the core dataset, the definition and measurement of title attractiveness, and the primary empirical strategies employed. Full technical details, data sources, and model implementation procedures are provided in Appendix A.1 and B.1.

Our dataset comprises 325,203 articles published between 2000 and 2022 in 328 economics journals included in the 2018 ABS Academic Journal Guide. We integrate article-level metadata from OpenAlex (Priem et al., 2022) with author-level information from Semantic Scholar, enabling us to control for key publication and author characteristics. This comprehensive dataset spans multiple tiers of journal quality and captures a wide spectrum of publication outcomes, citation patterns, and author attributes.

To enrich author-level variables, we match authors to Semantic Scholar records, obtaining publication counts, citation counts, and h-index values. We also estimate author gender using name-based predictions (Alexopoulos et al., 2023), recognizing the importance of controlling for potential gender-related differences in publication and citation dynamics (Bransch and Kvasnicka, 2022).

II.1 Measuring Title Attractiveness

Quantifying title attractiveness poses methodological challenges due to its subjective nature and large sample size. Simple observable title characteristics (e.g., length or punctuation) fail to capture deeper stylistic elements that influence editors, reviewers, and eventual citation patterns.

We employ a two-stage strategy. First, we conduct a stratified sampling of 1,900 titles from journals of varying quality. We recruit a panel of seven professors and seven students, spanning multiple economics subfields, to label each title as either "eye-catching" (1) or not (0). This annotator pool mitigates domain-specific biases and ensures a broadly representative judgment of title appeal.[2]

Second, we use these human-labeled data to fine-tune a BERT model (Devlin et al., 2018), an advanced NLP architecture well-suited to capturing nuanced linguistic features. By training the model on a representative set of human judgments, we achieve a scalable measure of title attractiveness for the entire dataset. BERT's ability to understand context, word order, and subtle stylistic elements provides a measure that is different from traditional dictionary-based or bag-of-words approaches.[3]

Initial descriptive analysis suggests that eye-catching titles are more prevalent in higher-ranked journals and are associated with greater citation counts. As illustrated in Figure 1, top-tier journals do not necessarily produce the largest absolute volume of attractive titles but feature a greater proportion of them. Turning to the relationship between title attractiveness and scholarly impact, Figure 2 indicates a positive association with citation counts; articles bearing more appealing titles generally accrue more references over time [4]. A temporal perspective emerges from Figure 3, which shows that as the academic environment has grown more competitive, the prevalence of eye-catching titles has increased, as has their share of total citations. Together, these patterns suggest that crafting more compelling titles may confer distinct advantages in garnering scholarly attention and influence.

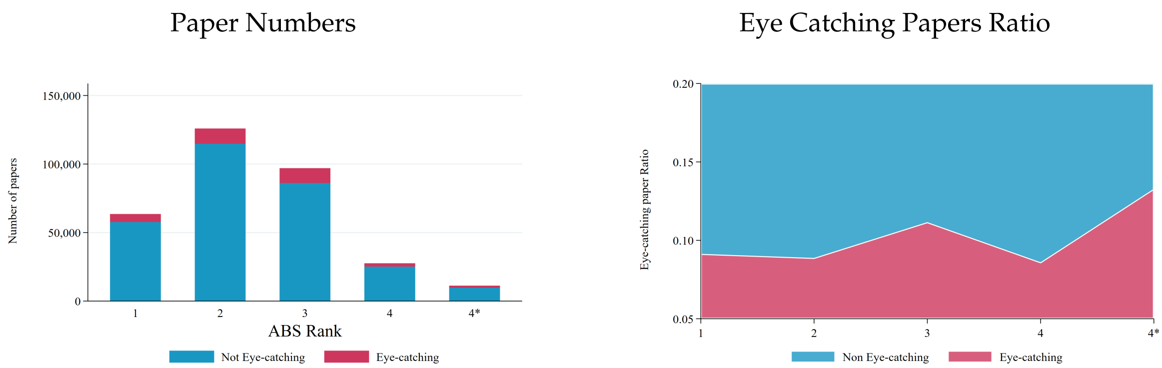
[2] We present a detailed description in Appendix B.1.

[3] We describe the fine-tuning of the BERT LLM model in Appendix B.1.

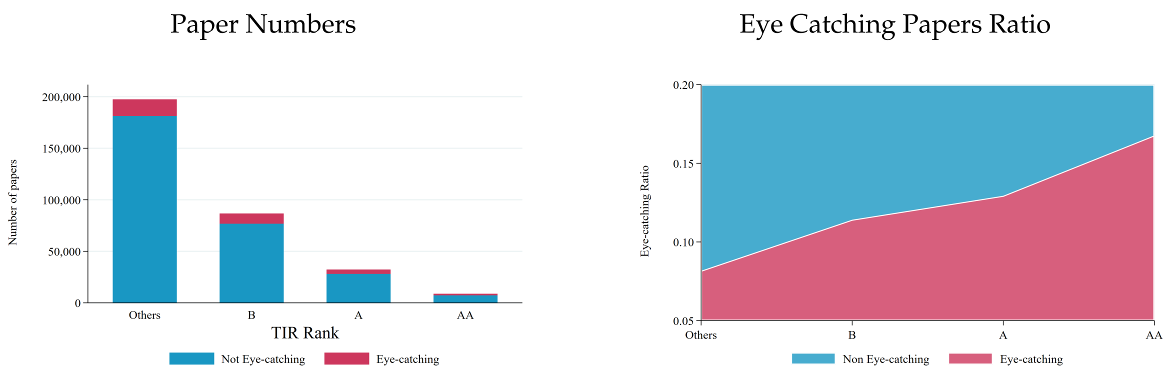
[4] The title attractiveness score is derived from our BERT model's prediction in the hidden layer, and the score of each title could be understood as a probability on a continuous scale from 0 to 1, where higher values indicate greater predicted attractiveness. Please refer to B.2 for more technical details.

FIGURE 1: Distribution and Proportion of Eye-catching Papers Across Journal Rankings

(A) ABS Journal Guide

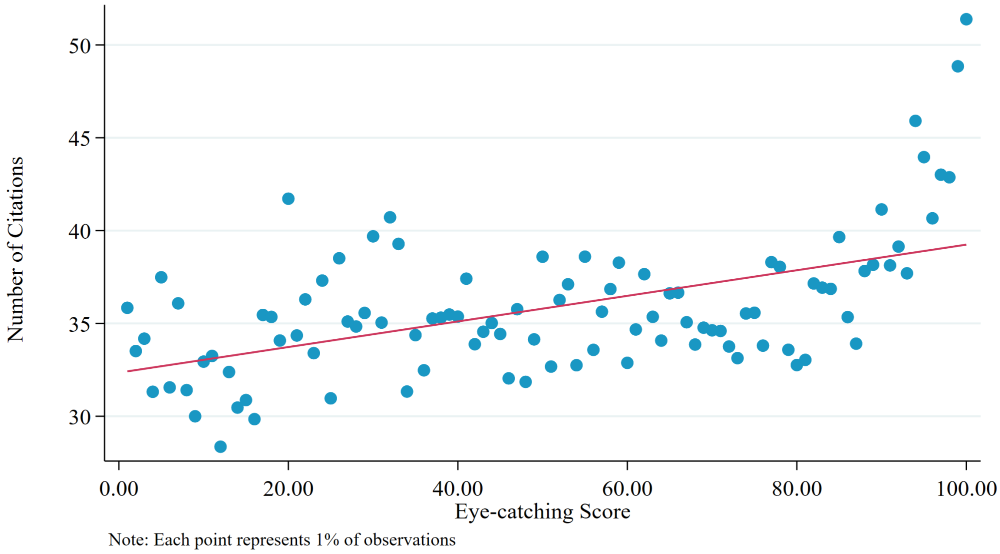


(B) Tinbergen Institute Ranking



<u>Notes:</u> Panel A uses the ABS Journal Guide ranking (1 to 4*, where 4* represents the highest quality), covering all economics journals listed in the 2018 ABS guide. Panel B uses the Tinbergen Institute Ranking, where "Others" refers to journals not classified as AA, A, or B in the Tinbergen list.

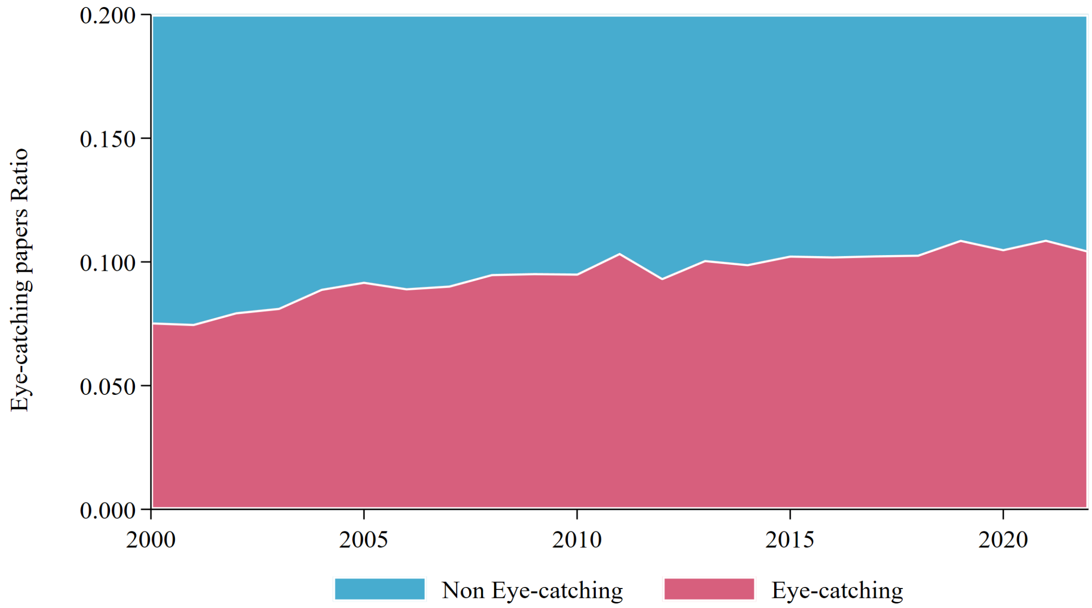
FIGURE 2: Relationship Between Title Attractiveness Score and Paper Citations



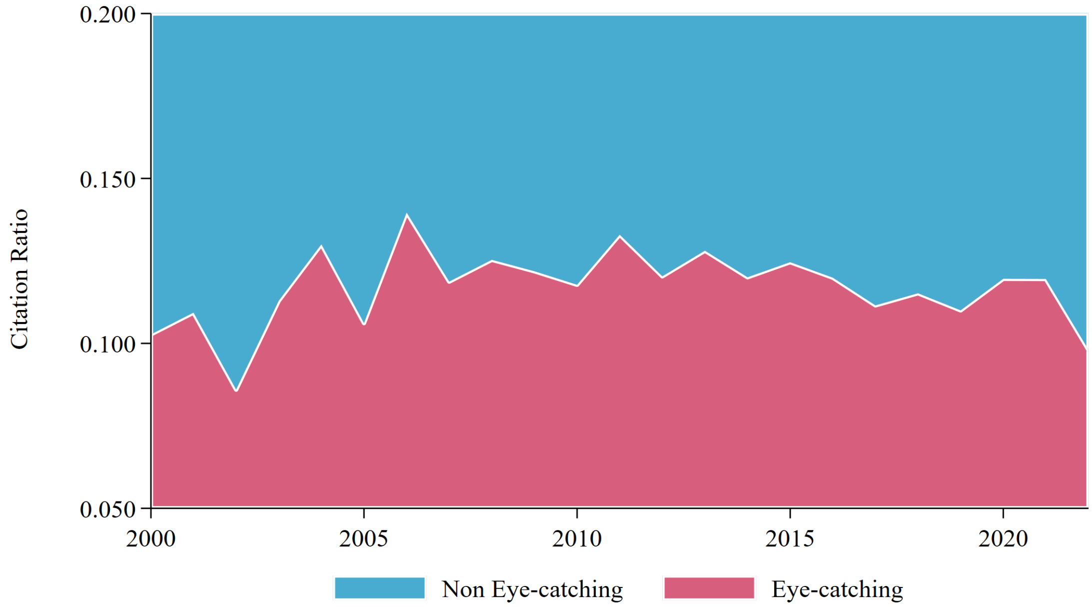
Notes: The figure presents a binned scatter plot of citation counts against percentile ranks of title attractiveness scores. Title Attractiveness scores from BERT model predictions are converted to percentile ranks (1-100) to better visualize the distribution, with each point representing 1% of observations. The red line represents the fitted linear relationship. Citation counts are adjusted for publication year to account for different exposure periods.

FIGURE 3: Year Trends in Eye-catching Papers' Distribution and Impact

(A) Paper Numbers Contributed by Eye Catching Papers



(B) Citations Contributed by Eye Catching Papers



Notes: The sample covers economics papers published between 2000 and 2022. Panel A shows the annual proportion of eye-catching papers among all publications. Panel B displays the share of total citations received by eye-catching papers in each year.

III. Empirical Specification

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

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

where ABS_i represents the ABS star rating of the journal in which article i is published. The key explanatory variable, Attractiveness_i, is a dummy indicator of whether the paper's title is considered attractive, as determined by the 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.[5]

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_t, the total citation count of the corresponding author in the year of publication 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 use 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%. Matching the corresponding author's affiliation with these rankings and creating 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 the baseline group. This granular approach allows me to capture the potential impact of institutional prestige on publication outcomes and citation counts, which may correlate with both title attractiveness and the dependent variables.

Finally, we include a vector of additional controls, X_i, which encompasses factors that may

[5] Note that in section V.4 we also consider a more granular control based on JEL journal. Further, we also consider interactions of such a JEL classification and the year of publication. The impact on the title attractiveness is similar.

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. We include yearly dummies Year_i of publication, to control for year-specific shocks in publication patterns.

Our second specification examines the relationship between title attractiveness and citation counts:[6]

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

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, γ_j. This allows me 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 1 presents the baseline estimation results for the impact of title attractiveness on publication prospects and citations. Columns (1) to (4) primarily examine the impact on journal quality. In Column (1), the coefficient for the 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. This finding aligns with the limited attention theory, suggesting that attractive titles may help papers capture the attention of editors and reviewers in an increasingly competitive publication landscape. In Column (2), we control for year-fixed effects of accounting for temporal variation in journal quality, and the coefficient remains stable. 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 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, the number of authors, and the number of female authors. We observe a slight increase in the coefficient for title attractiveness, which remains

[6] Another specification to see the impact of title attractiveness on the percent change of the citation using poisson estimation can be found in Appendix F.1

TABLE 1: 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			0.001**	-0.000			1.557***	1.414***	1.271***
			(0.001)	(0.001)			(0.154)	(0.159)	(0.156)
Team Citations			0.033***	0.032***			4.917***	4.856***	3.919***
			(0.001)	(0.001)			(0.218)	(0.222)	(0.219)
Theoretical			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 Dummies	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	325,203	325,203	301,121	285,803	325,203	325,203	301,121	285,803	286,792

Notes: The Table shows regression evidence for the impact of an eye-catching title on a publication's journal quality and citations. One observation corresponds to one individual article. Dependent variables are ABS journal ratings (columns 1-4) and total citations (columns 5-9). Eye-catching is a binary indicator based on BERT model classification. Correspond Citations and Team Citations are multiplied by 1000 for coefficient scaling. Theoretical and General are dummy variables indicating journal type based on manual classification. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings (top 1% to top 10% based on RePEc rankings). Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

stable and significant.

For the citation outcomes in Columns (5) - (9), the results similarly highlight the significant role of title attractiveness, potentially reflecting framing effects in how titles shape readers' perceptions and engagement with papers. In Column (5), the coefficient for the 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. 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 me 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. This persistent effect, even after controlling for journal quality, suggests that attractive titles may create lasting framing effects that influence how papers are perceived and cited within their respective fields.

V. ROBUSTNESS

In this section, we perform a series of robustness checks. These include employing alternative measures of journal quality and alternative characteristics of institutions and authors, incorporating JEL codes into the main specification, and adding additional control variables (title lengths and novelty index) that may influence the title attractiveness. [7]

V.1 Journal Quality

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 rank-

[7] Details on the inclusion of additional control variables and their potential effects on title attractiveness can be found in the Appendix C.1.

TABLE 2: Robustness Tests Using Alternative Journal Quality measures

	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	1.956***	0.000**	0.002***	-0.000
	(0.185)	(0.000)	(0.000)	(0.000)
Team Citations	3.991***	0.004***	0.002***	-0.002***
	(0.229)	(0.000)	(0.000)	(0.000)
Theoretical	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 Dummies	Yes	Yes	Yes	Yes
Affiliation Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	277,747	285,803	285,803	285,803

Notes: Column (1) uses RePEc rankings as a continuous measure of journal quality, with rankings inverted so that higher values indicate better journals for consistency with other measures. Columns (2)-(4) use binary indicators for journals classified as AA, A, or B in the Tinbergen Institute Ranking system. Eye-catching is a binary indicator based on BERT model classification. Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings (top 1% to top 10% based on RePEc rankings). Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

ings are in ascending order (i.e., rank 1 is the highest), there may be an issue of reversed interpretability in the 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 2 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 employs the Tinbergen Institute Journal Ranking, a well-established and authoritative metric in economics also used by Bramoullé and Ductor (2018) to measure publication quality. This ranking classifies journals into three categories: AA, A, and B. We treat these categories as binary variables in the regressions. Columns (2) to (4) of Table 2 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 finding aligns with predictions from the limited attention theory in a particularly illuminating way. 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, where most submissions achieve a high standard of quality (though not to the near-flawless level required for AA journals), the attention-capturing effect of an attractive title becomes particularly salient. In this intermediate quality tier, where papers compete intensely for limited journal space, the framing effect of an attractive title can be crucial in shaping editors' and reviewers' initial perceptions, potentially influencing their allocation of attention and ultimately affecting acceptance decisions. This pattern suggests that title attractiveness functions as a particularly effective attention-capturing mechanism in contexts where paper quality differences are relatively small and decision-makers face significant cognitive constraints in differentiating among high-quality submissions.

V.2 Institutions

TABLE 3: Robustness Tests Using Alternative Institution Characteristics

	Journal Quality					Citation				
	Ins.(Pub)	Ins.(Cite)	Ins.(AA)	Ins.(ABS)	Ins.FE	Ins.(Pub)	Ins.(Cite)	Ins.(AA)	Ins.(ABS)	Ins.FE
	(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	-0.007***	1.266***	1.268***	1.267***	1.267***	1.312***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.167)	(0.167)	(0.167)	(0.167)	(0.182)
Team Cit.	0.030***	0.025***	0.026***	0.026***	0.024***	4.136***	3.974***	4.105***	4.115***	4.028***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.237)	(0.235)	(0.237)	(0.237)	(0.249)
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. Control	0.446***	0.005***	0.822***	0.164***		-1.142*	0.046***	0.150	-0.079	
	(0.004)	(0.000)	(0.006)	(0.001)		(0.667)	(0.010)	(1.268)	(0.245)	
Year Dummies	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	251,976	251,976	251,976	251,976	245,696	252,933	252,933	252,933	252,933	246,617

Notes: Inst. (Pub) uses total publication counts, Inst. (Cite) uses total citation counts, Inst. (AA) uses number of AA-journal publications, and Inst. (ABS) uses cumulative ABS journal ratings to measure institutional quality. For each paper, institutional measures use the highest-ranked institution among all authors, multiplied by 1000 for coefficient scaling. Institution fixed effects are based on corresponding authors' affiliations. Corr. Cit. and Team Cit. are scaled by 1000. Eye Catch. is a binary indicator based on BERT model classification. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Theory and General are dummy variables for journal type. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We conduct a series of robustness checks to address potential concerns regarding institutional affiliations' impact on the main results. The baseline regression incorporated RePEc's institutional rankings (top 1% - top 10%) as a proxy for individual ability, complementing the controls for corresponding author citations and team citations. This approach addresses potential bias arising from the relationship between institutional prestige and researchers' capacity to craft attention-capturing titles, capturing author-related ability characteristics that may have been omitted from the initial specification.

While RePEc's rankings are widely recognized, they cover only 1,247 institutions, whereas the dataset encompasses 10,075 unique institutions. In the 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 the 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 the 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 the control variable.

Table 3 presents the results of these robustness checks. Columns (1)-(4) show the results using the newly developed institutional scores as control variables. We find that the coefficient on the measure of title attractiveness remains positive and significant at the 1% level across all specifications. This persistent effect under various institutional controls suggests that the attention-capturing mechanism of attractive titles operates independently of institutional prestige. 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, the 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, suggesting that the framing effects of attractive titles persist even after accounting for institutional quality. 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. The results continue to hold, demonstrating the robustness of the findings to this most stringent specification.

V.3 Authors

We further examine the robustness of the 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, using data from the Semantic Scholar database. Additionally, we adopt the most stringent approach by controlling for the

TABLE 4: Robustness Tests Using Alternative Author Characteristics

	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	0.031***	0.045***	0.011***	5.630***	6.115***	4.050***
	(0.001)	(0.001)	(0.001)	(0.250)	(0.225)	(0.315)
Theoretical	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 Dummies	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	285,803	285,803	299,847	286,792	286,792	300,897

Notes: Corresponding Hindex and Corresponding Paper represent the h-index and total publication count of corresponding authors from Semantic Scholar, respectively. Team Citations is scaled by 1000. Eye Catching is a binary indicator based on BERT model classification. Author FE indicates the corresponding author fixed effects. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings (top 1% to top 10% based on RePEc rankings). Theory and General are dummy variables for journal type. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

fixed effects of corresponding authors, considering individual effects for 113,632 authors in the sample. Table 4 presents the results of these 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 the title attractiveness variable remains robust and significant. This persistence suggests that the attention-capturing mechanism of attractive titles operates independently of author characteristics. 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 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.

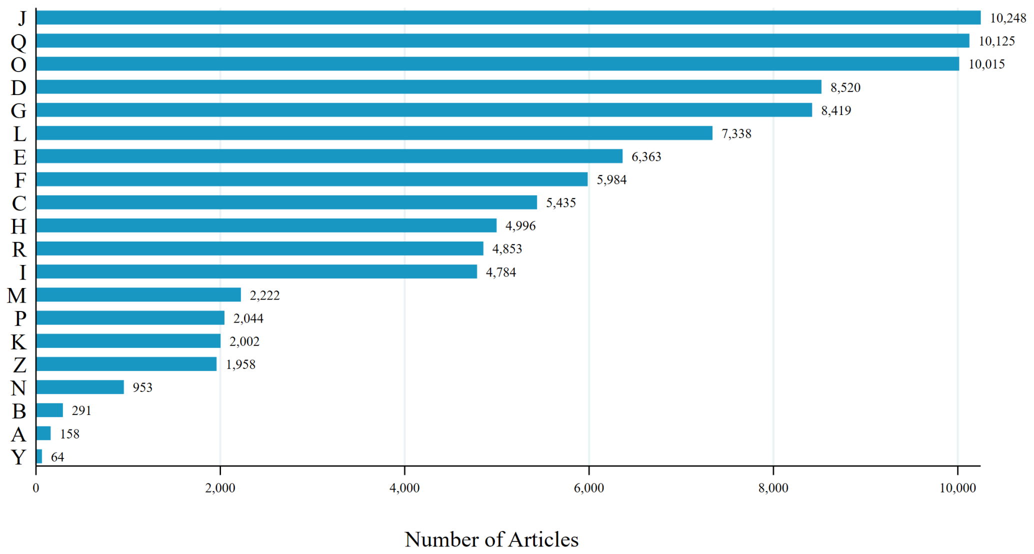
Columns (4) to (6) examine the impact of title attractiveness on citation counts. In Columns (4) and (5), we again alter the measure of author ability, but the regression results remain largely unaffected, and the findings continue to hold robustly. Interestingly, the coefficients for both the corresponding author's H-index and paper count are negative. This seemingly counter-intuitive 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 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, this does not fundamentally alter the baseline conclusions regarding the framing effects of title attractiveness on citations.

V.4 Including JEL

A potential concern with this analysis is that the trained model might exhibit systematic preferences for titles in specific research fields due to potential bias in the human annotators, potentially affecting the estimates. To address this concern, we introduce controls for research fields using JEL classification codes. Figure 4 presents the distribution of JEL codes in the sample.

FIGURE 4: Original JEL Distribution



Notes: Notes: Data based on 86,888 articles with available JEL codes from RePEc (approximately 26.7% of the total sample).

The initial approach involves collecting JEL codes directly from the RePEc database for the sample articles. However, this direct collection method yields JEL codes for only 86,888 articles, representing approximately 26.7% of the total sample. As shown in Figure 4, 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 the 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.[8]

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

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

predict metadata in economics research. The intuition behind this approach is that titles and abstracts typically contain the most salient information that researchers use to quickly assess a paper's field and relevance. 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, the DeBERTa model achieves strong predictive performance after training for 35,000 steps with a batch size of 4. The model attains an Jaccard index of 0.665 on the validation set, indicating that the predicted JEL codes overlap with actual codes by approximately two-thirds. This performance level suggests that the model effectively captures the field classifications.

TABLE 5: Robustness Estimation (JEL)

	Orginal JEL				Predicted JEL			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eye Catching	0.106***	0.103***	3.636**	3.614**	0.090***	0.089***	2.883***	2.941***
	(0.009)	(0.009)	(1.541)	(1.523)	(0.006)	(0.006)	(0.746)	(0.746)
Corr. Citations	0.003***	0.003***	0.871***	0.860***	0.001	0.001	1.284***	1.273***
	(0.001)	(0.001)	(0.306)	(0.301)	(0.001)	(0.001)	(0.152)	(0.152)
Team Citations	0.033***	0.032***	5.584***	5.479***	0.031***	0.031***	3.910***	3.896***
	(0.001)	(0.001)	(0.494)	(0.488)	(0.001)	(0.001)	(0.213)	(0.213)
Theoretical	0.509***	0.523***			0.534***	0.534***		
	(0.011)	(0.011)			(0.005)	(0.005)		
General	0.457***	0.451***			0.210***	0.209***		
	(0.008)	(0.008)			(0.004)	(0.004)		
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes	No	Yes	No	Yes
JEL Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
JEL × Year	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	86269	86269	86888	86888	297997	297997	298983	298983

Notes: Columns (1) - (4) use the JEL code directly collected from the RePEc Database, and Column (5) - (6) use the JEL code predicted by fine-tuned DeBertra. Columns (1),(2),(5), and (6) use journal quality (ABS rating) as the dependent variable. Columns (3),(4),(7), and (8) use citations as dependent variables. Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 presents regression results using both the original JEL codes from RePEc in Columns (1) - (2) and the predicted JEL codes from the 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 these results after controlling for JEL codes is particularly noteworthy. It demonstrates that the effects of title attractiveness persist even after accounting for field-specific variations in research presentation and evaluation. This suggests that the relationship between title attractiveness and academic success reflects a fundamental attention-capturing mechanism rather than field-specific preferences.

VI. DISCUSSION

In this section, we further investigate the heterogeneity effects of journal quality on citations. Besides, we compare the results derived from the Professor Model and the Student Model, examine the use of generative AI for title evaluation, utilize BERT logits as a measure of title attractiveness, and analyze the heterogeneity in citation patterns across different years.[9]

VI.1 Heterogeneity of Journal Quality

In this section, we examine how the impact of title attractiveness on citation counts varies across journals of different quality tiers, as classified by different ranking systems. Table 6 presents the results using the ABS Journal Guide. From Columns (1) to (5), we observe significant heterogeneity in the effect of title attractiveness across different journal tiers, revealing an inverted U-shaped relationship between title attractiveness and citation counts.

Specifically, for articles published in 1- or 2-star journals, title attractiveness shows 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 predictions based on limited attention theory. In lower-quality journals, readers' attention primarily focuses on paper quality rather than title appeal, reflecting a rational allocation of limited cognitive resources. If the paper's perceived quality or relevance is low, even an attention-capturing title fails to generate sustained interest or citations. For articles in higher-quality but not top-tier journals, title attractiveness becomes crucial in the attention allocation process. Given the substantial volume of quality research in these journals, readers face cognitive constraints in thoroughly examining every paper. Attractive titles serve as effective framing devices, helping readers efficiently allocate their limited attention to papers most relevant to their interests.

However, this attention-capturing mechanism becomes less critical for articles in top-tier jour-

[9]Details on the use of BERT logits as a measure of title attractiveness and the heterogeneity of citation patterns across years are provided in the Appendix D.1 and Appendix E.1.

TABLE 6: Hetergeneity: Heterogeneous Effects of Title Attractiveness Across Journal Tiers

	ABS Ranking					Tinbergen Institute Ranking			
	ABS:1	ABS:2	ABS:3	ABS:4	ABS:4*	AA	A	B	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eye Catching	-0.0373	-0.675	3.710***	13.75*	3.137	3.148	8.325**	2.861***	0.240
	(0.390)	(0.505)	(1.033)	(5.735)	(9.600)	(9.599)	(2.839)	(0.759)	(0.699)
Corr. Citations	0.440***	0.758***	1.393***	2.095***	0.385	0.386	1.225**	1.045***	1.276***
	(0.0948)	(0.117)	(0.168)	(0.505)	(1.088)	(1.088)	(0.404)	(0.178)	(0.177)
Team Citations	0.681***	1.675***	3.516***	5.916***	10.23***	10.23***	6.100***	3.220***	2.038***
	(0.148)	(0.148)	(0.301)	(0.708)	(1.447)	(1.447)	(0.660)	(0.320)	(0.210)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	57,295	116,341	92,955	26,437	7,967	7,965	34,682	80,990	177,358

Notes: Each column presents results for papers published in journals of different ABS ranking tiers (1 to 4*, where 4* represents the highest quality). The dependent variable is citation count. Eye Catching is a binary indicator based on BERT model classification. Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings (top 1% to top 10% based on RePEc rankings). Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

nals (top 5) for several reasons. First, the prestigious status of these journals itself serves as a powerful framing device, creating a default assumption of quality and importance. Second, top journals typically publish relatively few papers, significantly reducing the cognitive burden on readers. With a manageable volume of articles to track, scholars can allocate sufficient attention to each paper without relying heavily on title attractiveness as a screening mechanism. Moreover, given the high stakes of missing important contributions in top journals, researchers are likely to examine all articles in these outlets regardless of their titles. This creates a context where the marginal benefit of an attractive title is minimized, as both the journal's prestige and its limited publication volume already ensure comprehensive attention from the academic community.

We also conducted a similar heterogeneity analysis using the Tinbergen Institute Rank, with results presented in Columns (6) - (9). The findings remain consistent with those from the ABS Journal Guide analysis. Specifically, the effect of title attractiveness on citations concentrates in journals ranked as A and B, which is consistent with our results using ABS Journal Guide.

VI.2 Professor Preference and Student Preference

TABLE 7: Comparing Professor and Student Model

	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	1.275***	-0.000	1.275***	-0.000	1.275***	-0.000	1.275***
	(0.001)	(0.151)	(0.001)	(0.151)	(0.001)	(0.151)	(0.001)	(0.151)
Team Citations	0.032***	3.877***	0.032***	3.876***	0.032***	3.878***	0.032***	3.878***
	(0.001)	(0.211)	(0.001)	(0.211)	(0.001)	(0.211)	(0.001)	(0.211)
Theoretical	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)	
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes	No	Yes	No	Yes
cmidrule(lr)1-9 Affiliation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	300,006	300,995	300,006	300,995	300,006	300,995	300,006	300,995

Notes: Results compare BERT models trained separately on professor annotations (columns 1-4) and student annotations (columns 5-8). Columns (1), (3), (5), and (7) use journal quality (ABS rating) as the dependent variable, and columns (2), (4), (6), and (8) use citations. Eye Catching Logits uses raw prediction scores, while Eye Catching uses binary classification with 0.5 threshold. Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The perception of title attractiveness may vary between established scholars and early-career researchers, potentially reflecting different attention allocation patterns and framing preferences. To investigate this possibility and provide an additional robustness check for the main results, we conducted separate analyses using BERT models trained on professors' and students' evaluations of title attractiveness. Table 7 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 [10] as

[10]The logits scores is derived from our BERT model's prediction in the hidden layer, and the score of each title could be understood as a probability on a continuous scale from 0 to 1, where higher values indicate greater predicted attractiveness. Please refer to B.2 for more technical details.

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 become 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 attractive 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.3 Using Generative AI in Title Evaluation

TABLE 8: Title Attractiveness Assessed by Different Large Language Model

	CHATGPT 3.5		LLAMA3-8B		GPT-4O	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Normal Prompt						
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	-0.002	1.082***	-0.002	1.088***	-0.002	1.085***
	(0.002)	(0.375)	(0.002)	(0.375)	(0.002)	(0.375)
Team Citations	0.032***	3.308***	0.032***	3.307***	0.032***	3.306***
	(0.002)	(0.480)	(0.002)	(0.480)	(0.002)	(0.480)
Theoretical	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)	
Observations	30,149	30,247	30,149	30,247	30,149	30,247
Panel B: Economists Prompt						
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	-0.002	1.088***	-0.002	1.089***	-0.002	1.084***
	(0.002)	(0.374)	(0.002)	(0.375)	(0.002)	(0.375)
Team Citations	0.032***	3.314***	0.032***	3.306***	0.032***	3.308***
	(0.002)	(0.480)	(0.002)	(0.480)	(0.002)	(0.480)
Theoretical	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)	
Observations	30,149	30,247	30,149	30,247	30,149	30,247
Year Dummies	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

Notes: Results are based on a 10% random sample of the full dataset. Title attractiveness is evaluated using three LLMs with standard prompts similar to those given to human annotators. Columns (1), (3) and (5) use journal quality (ABS rating) as dependent variable, columns (2), (4) and (6) use citations. Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

To further validate these 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 the 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 the human-annotated sample, providing a supplementary check to mitigate potential biases in the 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. This 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 the total dataset. While this limitation reduces statistical power, it allows me to maintain cost-effectiveness while still providing valuable insights.

Table 8 present the results of this analysis using GenAI models to evaluate title attractiveness. In Panel A in Table 8, we use a prompt similar to that given to the human annotators, while Panel B employs a prompt that instructs the AI to assume the perspective of an experienced economist who has read many papers.

In Panel A, 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 in Columns (1) and (5), the coefficients are 0.066 and 0.121 respectively. For citations in Columns (2) and (6), we find 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 suggests that the aggregate preference represented by LLAMA3-8B may be less effective at identifying titles that capture broad reader attention.

A notable trade-off emerges between publication prospects and citation rates of eye-catching titles. While the titles GPT-4o identifies as eye-catching have higher publication prospects, they demonstrate relatively lower citation rates. This pattern may reflect fundamental differences in attention allocation between editors/reviewers and potential citers. Editors and reviewers, as carefully selected field experts, are attuned to field-specific preferences and standards. In contrast, citing authors come from diverse fields and operate under different cognitive constraints. Consequently, titles that effectively capture the attention of seasoned scholars in a specific field may not resonate as strongly with the broader academic audience, leading to the observed trade-off between publication success and citation impact. This pattern parallels the citation dynamics of top 5 journals, where some papers, despite their prestigious publication venue, accumulate fewer citations than articles in more specialized journals.

Given these patterns, we explored whether prompting the model to adopt the perspective of an experienced economist might align the model's preferences more closely with those of high-level economics editors and reviewers. In Panel B, which employs the economist-framed prompt, reveals similar patterns but with notable variations. For ChatGPT 3.5, the economist prompt yields a smaller coefficient for journal quality (0.044 vs 0.066) but a larger coefficient for citations (7.036 vs 5.947). LLAMA3-8B demonstrates enhanced performance with the economist prompt, showing a larger coefficient for journal quality (0.098 vs 0.071) and, notably, a significant effect on citations (2.706) that was absent with the neutral prompt. GPT-4 exhibits more consistent results across prompts, though with slightly lower coefficients for the economist prompt.

Contrary to initial expectations, the economist prompt did not consistently align the models' preferences with 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, demonstrated increased coefficients. This pattern suggests that the prompt may have inadvertently emphasized the model's perspective as a reader rather than an editor or reviewer. These results underscore the challenges and inherent unpredictability of using prompt engineering to fine-tune model preferences, indicating that careful consideration is essential when attempting to align AI models with specific human preferences.

These findings provide robust support for the main results. The consistency of positive and significant effects across different GenAI models suggests that the results reflect genuine patterns rather than artifacts of specific human biases or limited sample sizes. Moreover, the statistical significance across most models demonstrates the potential feasibility of using GenAi as a substitute for human subjects in preference experiments of this nature. However, the results also highlight important caveats regarding the use of prompt engineering to approximate specific types of human preferences, emphasizing the necessity for careful and rigorous testing in such applications.

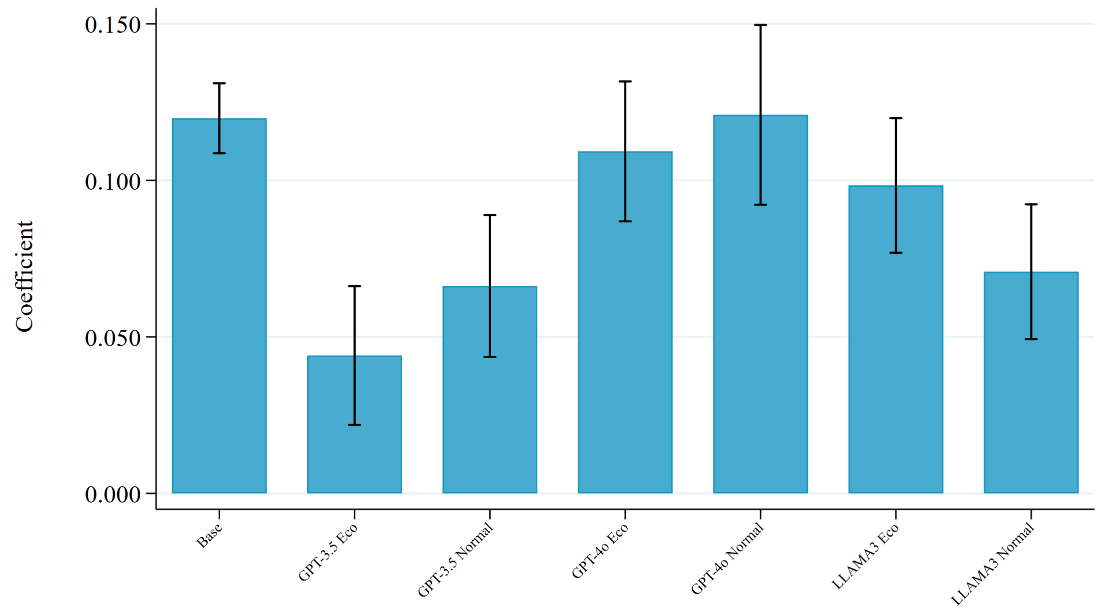
The observed differences also illuminate the nuanced nature of title attractiveness, suggesting that various aspects of attractiveness may carry different weights depending on the outcome of interest (journal quality versus citations) and the specific AI model employed. This insight offers practical value for researchers seeking to optimize their titles for specific academic goals and in selecting appropriate AI tools for analysis. The varying performance of different models and prompts suggests that multiple approaches might be necessary to comprehensively assess title attractiveness across different academic contexts.

To provide a more comprehensive comparison, Figure 5 presents a visual representation of the coefficients from the main human-annotated model alongside those from the various GenAI models.

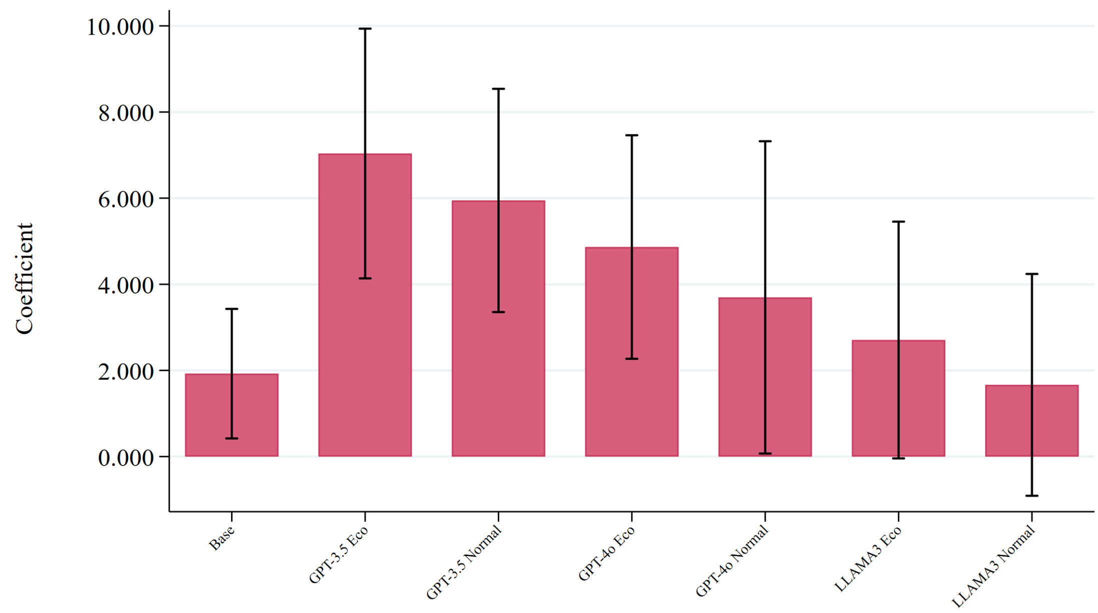
Panel A of Figure 5 compares the coefficients for journal quality across different models. We find that GPT-4's estimated coefficients are closest to those of the BERT model fine-tuned on human samples, suggesting that more capable models have preference recognition abilities in specialized fields that approach those of human experts. Conversely, ChatGPT 3.5 shows relatively weaker performance, implying that its preferences align more closely with general

FIGURE 5: Comparison of Title Attractiveness Effects Across Different Models

(A) ABS Journal Guide



(B) Citations



Notes: Coefficients from regressions using different model assessments of title attractiveness. ”Bert” refers to the BERT model trained on human annotations. For LLMs (GPT-3.5, GPT-4, LLAMA3), ”Normal” indicates standard prompts and ”Econ” indicates economist-perspective prompts. Panel A shows effects on journal quality (ABS rating); Panel B shows effects on citations. Vertical bars represent 95% confidence intervals. All estimates based on a 10% random sample of the full dataset with the same control variables as baseline specifications.

interests rather than specialized academic judgments. This pattern is consistent with its larger coefficient for citations. LLAMA-3 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 the human-annotated model. This suggests that ChatGPT 3.5 may be particularly adept at identifying title characteristics that correlate with higher citation rates. GPT-4, while performing well, shows a more conservative estimate compared to the human-annotated model. LLAMA-3 again falls between the two, but with a notably lower coefficient compared to its performance on journal quality.

These observations highlight an interesting dichotoour: while more advanced models like GPT-4 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 pattern could reflect the fundamental difference between these two outcomes - publication quality being more dependent on expert assessment, while citations may be influenced by broader appeal factors that ChatGPT 3.5 captures effectively.

The proximity of coefficients between GenAI models and the 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 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-4 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 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, consistent with limited attention theory, suggests that title attractiveness serves as a crucial attention-capturing mechanism where cognitive constraints are most binding - in the competitive middle segment of academic publishing. At this level, papers must actively compete for reader attention, while top-tier journals' inherent prestige and low publication volume naturally command attention.

Third, the 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 the 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 academic publishing. For individual researchers, the results suggest that investing time in crafting attention-capturing titles can yield meaningful returns in terms of publication success and scholarly impact. For journal editors and reviewers, the findings highlight the need to consider how attention mechanisms and framing effects might influence evaluation processes. For the broader academic community, this work underscores the evolving nature of academic communication in an environment increasingly characterized by competition for limited attention.

This study also points to several promising directions for future research. First, investigating whether similar attention-capturing and framing mechanisms exist in other disciplines could illuminate field-specific differences in how title characteristics influence academic success. Second, exploring how title attractiveness interacts with other attention-directing features, 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 cognitive mechanisms represents an important area for future investigation.

References

Alexopoulos, M., Lyons, K., Mahetaji, K., Barnes, M. E., and Gutwillinger, R. (2023). Gender inference: can chatgpt outperform common commercial tools? In *Proceedings of the 33rd Annual International Conference on Computer Science and Software Engineering*, pages 161–166.

Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15(1):659–688.

Bramoullé, Y. and Ductor, L. (2018). length. *Journal of Economic Behavior & Organization*, 150:311–324.

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

Card, D. and DellaVigna, S. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1):144–161.

Chan, H. F., Önder, A. S., Schweitzer, S., and Torgler, B. (2023). Twitter and citations. *Economics Letters*, 231:111270.

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Chen, Y., Liu, T. X., Shan, Y., and Zhong, S. (2023). The emergence of economic rationality of gpt. *Proceedings of the National Academy of Sciences*, 120(51):e2316205120.

Cohn, J. B., Liu, Z., and Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2):529–551.

Dell, M. (2024). Deep learning for economists. Working Paper 32768, National Bureau of Economic Research.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ductor, L. and Visser, B. (2022). When a coauthor joins an editorial board. *Journal of Economic Behavior & Organization*, 200:576–595.

Feld, J., Lines, C., and Ross, L. (2024). Writing matters. *Journal of Economic Behavior & Organization*, 217:378–397.

Galiani, S., Gálvez, R. H., and Nachman, I. (2023). Unveiling specialization trends in economics research: A large-scale study using natural language processing and citation analysis. Technical report, National Bureau of Economic Research.

Gnewuch, M. and Wohlrabe, K. (2017). Title characteristics and citations in economics. *Scientometrics*, 110:1573–1578.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2023). The voice of monetary policy. *American Economic Review*, 113(2):548–584.

Guo, F., Ma, C., Shi, Q., and Zong, Q. (2018). Succinct effect or informative effect: The relationship between title length and the number of citations. *Scientometrics*, 116:1531–1539.

Hadavand, A., Hamermesh, D. S., and Wilson, W. W. (2024). Publishing economics: How slow? why slow? is slow productive? how to fix slow? *Journal of Economic Literature*, 62(1):269–293.

Hamermesh, D. S. (2018). Citations in economics: Measurement, uses, and impacts. *Journal of Economic Literature*, 56(1):115–156.

Hansen, S., Lambert, P. J., Bloom, N., Davis, S. J., Sadun, R., and Taska, B. (2023). Remote work across jobs, companies, and space. Technical report, National Bureau of Economic Research.

Hasan, S. A. and Breunig, R. V. (2021). Article length and citation outcomes. *Scientometrics*, 126:7583 – 7608.

Heckman, J. J. and Moktan, S. (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, 58(2):419–470.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research.

Hotbllino, H. (1929). Stability in competition. *The Economic Journal*, 39(153):41–57.

Jelveh, Z., Kogut, B., and Naidu, S. (2024). Political language in economics. *The Economic Journal*, page ueae026.

Kinney, R., Anastasiades, C., Authur, R., Beltagy, I., Bragg, J., Buraczynski, A., Cachola, I., Candra, S., Chandrasekhar, Y., Cohan, A., et al. (2023). The semantic scholar open data platform. *arXiv preprint arXiv:2301.10140*.

Kovács, B., Hsu, G., and Sharkey, A. (2024). The stickiness of category labels: Audience perception and evaluation of producer repositioning in creative markets. *Management Science*, 70(9):6315–6335.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. (2020). Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Ma, C., Li, Y., Guo, F., and Si, K. (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.

Priem, J., Piwowar, H., and Orr, R. (2022). Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2):221–243.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sun, C., Qiu, X., Xu, Y., and Huang, X. (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*, pages 194–206. Springer.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.

Walker, J. T., Fenton, E., Salter, A., and Salandra, R. (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(3):730–747.

Zhang, Z., Yang, K., Zhang, J. Z., and Palmatier, R. W. (2023). Uncovering synergy and dysergy in consumer reviews: A machine learning approach. *Management Science*, 69(4):2339–2360.

Online Appendix

A.1. Data Construction and Methodological Framework

This section describes the construction of our dataset and methodological framework for investigating the impact of title attractiveness on publication outcomes and citation rates in economics. We detail the data collection process, the construction of key variables, the development of a novel approach to measure title attractiveness, and employ some methodological approaches to address potential measurement challenges in assessing title attractiveness.

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 the 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 1,900 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, enabling scalable and consistent assessment of title attractiveness across the entire corpus.

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

A.1 Data Description

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.

A.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 et al., 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). [11] 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.

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 me 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 the estimates.

The AJG ratings serve a dual purpose in this 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 the study period.

A.1.2 OpenAlex

The primary source of bibliometric data for this study is the OpenAlex database, an open-source, comprehensive index of scholarly works, authors, venues, institutions, and concepts (Priem et al., 2022). This resource provides a wealth of bibliometric data crucial for the 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) [12].

The 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 the results, given the breadth and depth of the dataset.

To ensure the integrity and relevance of the sample, we implemented a rigorous data cleaning

[11] The choice of this particular edition is predicated on the observation that economics journal rankings have exhibited relative stability across recent editions.

[12] All these data are retrieved from OpenAlex API in January, 2024. This implies the citation data are counted before the Jan, 2024.

protocol. [13] Post-cleaning, the final dataset comprises 325,203 unique articles, representing a robust and comprehensive sample of economics literature spanning two decades.

A.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 the 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 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 et al., 2023).

For this study, we extracted information on 210,420 unique authors identified in the OpenAlex literature dataset and obtained three key metrics for each author: total citation count, h-index, and publication count from Semantic Scholar.[14] 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 pretrained models in gender prediction tasks (Alexopoulos et al., 2023). By incorporating this gender data, we are able to control for potential gender-related effects in the analysis of title attractiveness and publication outcomes.

[13] 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

[14] To mitigate the risk of homonyour, we implemented a rigorous matching process, ensuring that each author in the sample had at least one publication in Semantic Scholar corresponding to a publication in the OpenAlex dataset. This approach enhances the reliability of the author-level data.

B.1. Details on Measuring Title Attractiveness

B.1 Overview

In investigating 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.[15] The second challenge arises from the scale of our dataset, comprising 325,203 article titles, which makes comprehensive human evaluation prohibitively costly.

To address these challenges, we turn to machine learning techniques, which have proven effective in analyzing large-scale unstructured data in economics.[16] While traditional dictionary methods are commonly used for concept detection (Ash and Hansen, 2023), the subjective and context-dependent nature of title attractiveness makes such approaches inadequate.[17] For instance, unlike financial sentiment analysis where predefined dictionaries can be effectively employed (Loughran and McDonald, 2011), capturing what makes a title ”eye-catching” requires understanding complex linguistic patterns and contextual relationships.

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 from diverse economic subfields, 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 with bag-of-words representation - for example, Jelveh et al. (2024) used this approach with 98,479 phrases from 20,029 economics papers to predict paper ideology. However, in the context of assessing title attractiveness, this approach faces two significant limitations. First, attractive and unattractive titles often have high word overlap, meaning they appear similar in the embedding space despite differing in attractiveness (Dell, 2024). Second, bag-of-words models struggle to capture the nuanced use of words and their contextual relationships, which are critical in judging 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”

[15] 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.

[16] Machine learning approaches excel at extracting meaningful information from high-dimensional data and transforming it into lower-dimensional representations (Ash and Hansen, 2023).

[17] A detailed discussion of why dictionary-based methods are unsuitable for assessing title attractiveness, including examples and technical limitations, is provided in Appendix B.3.1

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. Despite sharing key words such as ”time”, ”death”, ”health care”, ”expenditure”, and ”predictions”, the first title is noticeably more attractive. The phrase ”Time to include time to death?” juxtaposes the concept of time in two different contexts, creating an intriguing wordplay, while ”The future of health care expenditure predictions” adds a forward-looking spin to 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 would be lost in a bag-of-words representation.

Given the complexity of this task, we employ BERT (Bidirectional Encoder Representations from Transformers), one of the most advanced natural language processing (NLP) models, to address the research question.[18] 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, endowing BERT with robust language understanding capabilities crucial for detecting subtle linguistic features in titles.[19]

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 the 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, we can analogize it to the human education process. [20] Attempting to train a neural network model from scratch to judge whether economics titles are eye-catching would be akin to teaching a newborn baby with no prior knowledge. Even with substantial time investment, the child might 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

[18]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

[19]For example, given the sentence ”The [MASK] brown fox jumps over the [MASK] dog,” BERT attempts to predict the masked words during the pre-training process. This training approach mimics human language comprehension processes

[20]More details of BERT and idea of transfer learning can be referred to Dell (2024)

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 fine-tuning of BERT involves adding a linear classifier to the pre-trained model while keeping the original parameters fixed (Dell, 2024). This transfer learning paradigm has become increasingly prevalent in frontier research across computer science and management science disciplines.[21] Despite the relatively limited application of BERT in economic research thus far, BERT shows significant potential in two key dimensions. First, it provides a sophisticated means to incorporate unstructured textual data into empirical research, capable of mapping and extracting low-dimensional vectors such as categories, topics, or sentiments from complex textual data (Ash and Hansen, 2023). This capability substantially enriches the data sources available for economic research, allowing for the analysis of previously untapped information. Second, these models are particularly effective for large-scale annotation tasks, where researchers can efficiently process extensive datasets using models fine-tuned on relatively small annotated samples (Dell, 2024).[22]

In recent years, BERT has been employed in several pioneering and influential studies in economics. The earliest application of BERT in economics appears to be in Shapiro et al. (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 original 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.[23] Their fine-tuned BERT model achieved over 80% accuracy in predicting hawkish, neutral, and dovish stances of FOMC statements.

---

[21] Detailed technical implementation of BERT fine-tuning and comprehensive examples of its applications across different domains are provided in Appendix B.3.2

[22] 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 (300-1,000 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.

[23] 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)

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 "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 this 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 me 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 the 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 1,900 papers. This diverse pool of annotators helps to mitigate potential biases and ensures that the trained BERT model captures a broad consensus on what constitutes an attractive title in economics.

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

B.1.1 Survey Design and Implementation

The initial phase of the 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 1,900 papers from the total sample, ensuring proportional representation across journal quality tiers and mitigating potential biases in the 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."

The 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 the 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.[24]

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.

The 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 the 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 B.1 shows the detail of the annotator pool, including their type (student or professor) and their research field.

In total, we collected 1,900 title annotations from the diverse pool of annotators. Among these annotations, 266 titles were identified as "eye-catching," suggesting that the annotators maintained relatively high standards in their evaluations. Interestingly, we observed substantial heterogeneity in individual annotation patterns. The most conservative evaluation came from a professor who identified only 3 titles in their 100-title sample as eye-catching, while the most liberal assessment came from a student who classified 38 titles in their sample as eye-catching. This variation in annotation patterns across academic ranks and experience levels underscores the subjective nature of title attractiveness and validates the strategy of incorporating diverse perspectives in the training data.

The diversity in the training data is particularly valuable for the 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.

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

TABLE B.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

B.1.2 BERT Model Fine-tuning

After obtaining the annotated samples, we proceeded to fine-tune the BERT model using the bert-base model from Hugging Face.[25] The fine-tuning process required careful consideration of data distribution to ensure robust model performance. To address potential biases in the training data, we implemented strategies to balance both the distribution of positive versus negative samples and the contribution of different annotators.[26] The final training dataset comprised 2,800 samples, with 20% reserved for validation. The model achieved an accuracy of 0.86 on the validation set, comparable to inter-annotator agreement levels.

A crucial consideration in evaluating the model's performance is understanding the inherently subjective nature of the task. Unlike traditional classification problems where ground truth exists, title attractiveness involves substantial individual variation in preferences. In this context, the model's performance should be assessed not against an absolute standard, but rather its ability to capture consensus among diverse annotators.

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, the 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 individ-

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

[26]The specific methodological approaches for handling data imbalances are discussed in detail in Appendix B.4

uals 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.

The 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.

To validate the 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% on the validation set - compared to the final model trained on all 14 annotators' data (86% accuracy). 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.

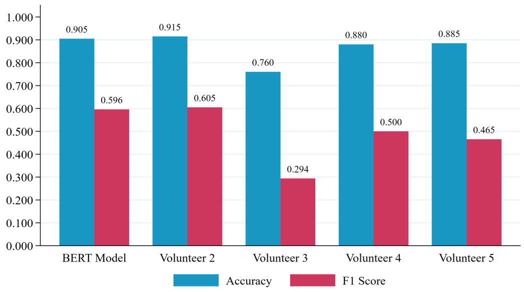
To further investigate the model's ability to capture collective preferences, we conducted a comparative analysis where these five annotators and the fine-tuned BERT model evaluated the same set of 200 paper titles. In Figure B.1, we present pairwise comparisons of accuracy and F1 scores between each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

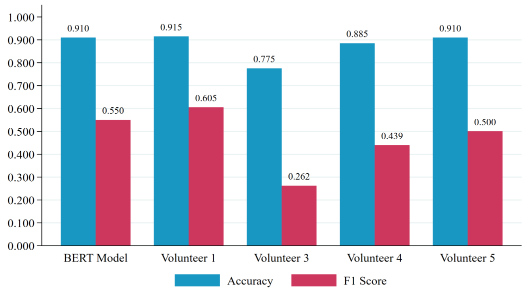
Second, and more importantly, the fine-tuned BERT model demonstrates remarkable consistency with human judgments across all comparisons. When compared against annotators 1, 2, and 3, the BERT model achieves the second-highest agreement scores among all pairwise comparisons. Moreover, it shows the highest 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

FIGURE B.1: Comparison of BERT Model Performance with Individual Human Annotators

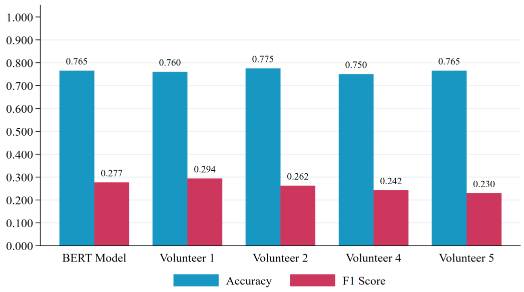
(A) Compared with Volunteer 1



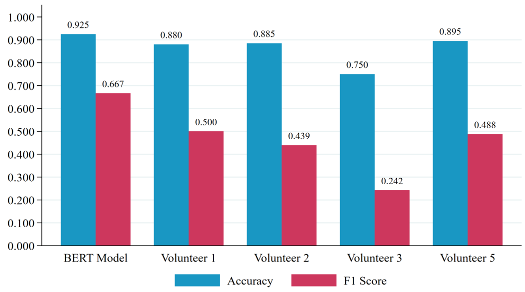
(B) Compared with Volunteer 2



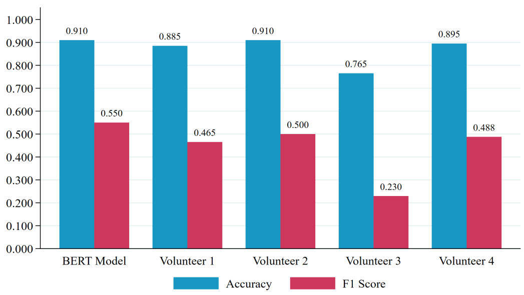
(C) Compared with Volunteer 3



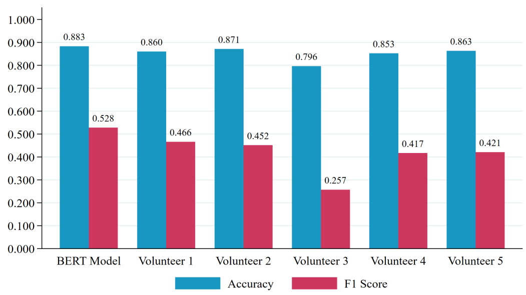
(D) Compared with Volunteer 4



(E) Compared with Volunteer 5



(F) Average score between BERT Model and Other Volunteer



Notes: Each panel compares the BERT model's performance against one annotator as reference, showing agreement levels with other annotators. The BERT model demonstrates consistently higher or comparable performance relative to human annotators, achieving the best average agreement scores (accuracy: 0.883, F1: 0.528) across all pairwise comparisons.

synthesize a composite preference that better represents the collective judgment of the group.

This analysis provides compelling evidence that the 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.

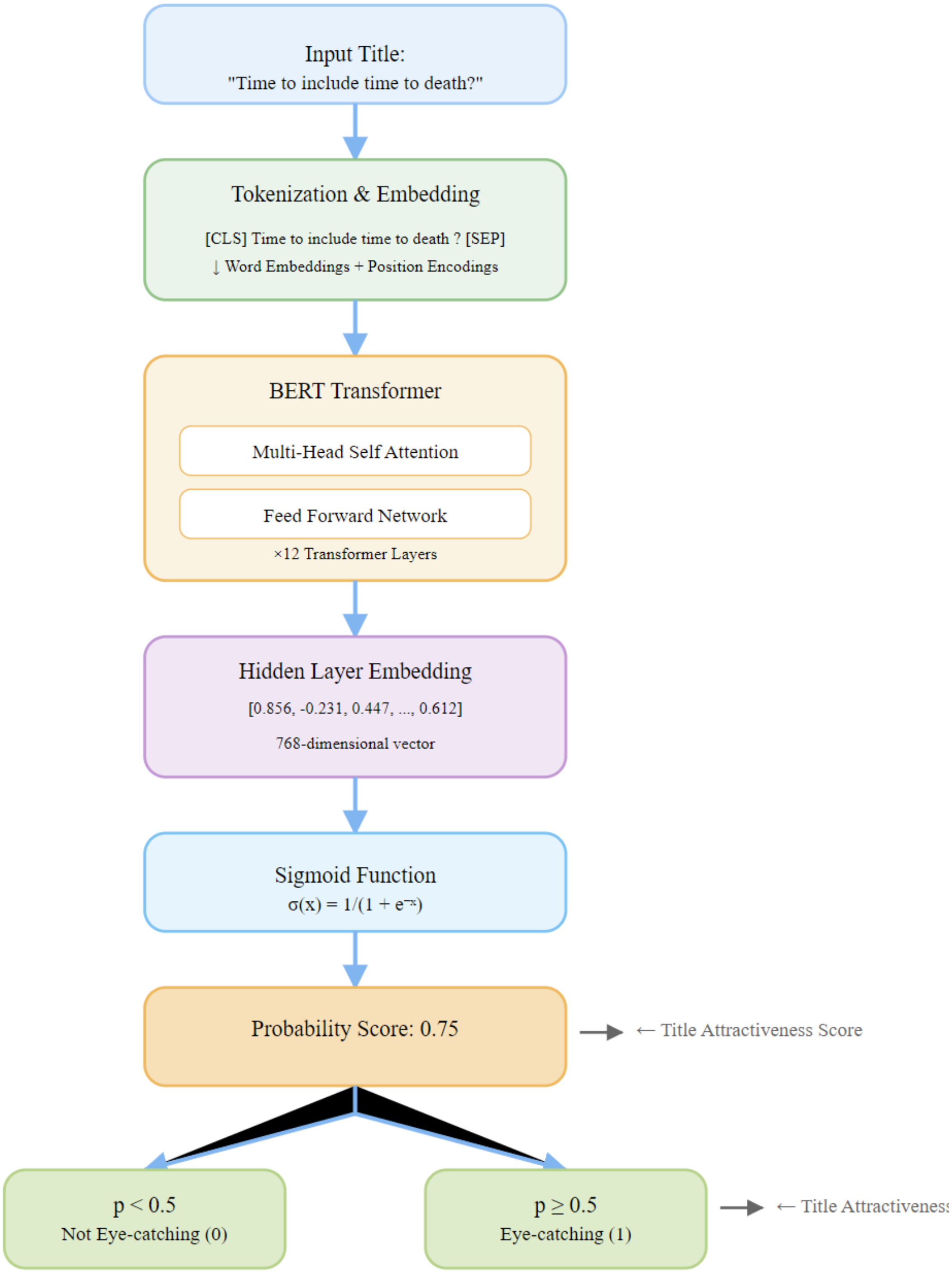
B.2 Details on BERT Outputs

BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model that processes text through multiple transformer layers to generate contextual representations. When fine-tuned for binary classification tasks like title attractiveness assessment, BERT's architecture processes titles through several stages to generate the final classification:

1. **Input Processing:**
   - Each title is first tokenized into word pieces using BERT's vocabulary
   - Special tokens [CLS] and [SEP] are added at the start and end
   - Tokens are converted to embeddings and positional information is added
2. **Transformer Layers:**
   - The embeddings pass through 12 transformer layers (in BERT-base)
   - Each layer applies self-attention mechanisms and feed-forward networks
   - The [CLS] token accumulates information about the entire title
3. **Classification Process:**
   - The final [CLS] token representation is processed through a hidden layer, generating a probability score between 0 and 1
   - This probability score is then converted to a binary classification (0 or 1) using a threshold of 0.5
   - The binary output represents BERT's final judgment of whether a title is eye-catching

These process are showed in the Figure B.2 for better understanding. In the main analysis, I primarily use BERT's final binary output (0 or 1) as the measure of title attractiveness. This choice is motivated by several considerations. First, this binary classification directly corresponds to our training data where human annotators made binary judgments. Second, it provides a more intuitive interpretation compared to probability scores, especially given that the training data itself is binary. Third, this approach maintains consistency with our subsequent analyses using other Large Language Models (LLMs), where only binary classifications are available due to limited access to internal model representations, particularly in closed-source models.

FIGURE B.2: BERT Processing Pipeline and Output Generation



While the intermediate probability scores (ranging from 0 to 1) from BERT's penultimate layer carry meaningful information about the degree of title attractiveness, I use them only in specific supplementary analyses. These include visualization of the relationship between title attractiveness and citation patterns (Figure 2) and additional robustness checks (Tables D.1 and 7) .[27]

B.3 Details on Measuring the Title Attractiveness

B.3.1 Why dictionary-based method is not suitable for measuring title attractiveness

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 personal rules or perceptions, we would inevitably introduce individual 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.

B.3.2 BERT fine-tuning in Computer Science and Management Science Research

The transfer learning paradigm 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 et al., 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 et al., 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 has also become very popular in frontier research in recent years. Kovács et al. (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 studies find that fine-tuned BERT models achieve great performance on their tasks and highly outperform existing models in their respective domains.

[27] For a comprehensive technical overview of BERT's architecture, see Devlin et al. (2018).

B.4 Details on BERT Model Fine-tuning

B.4.1 Techical Details and Training Result

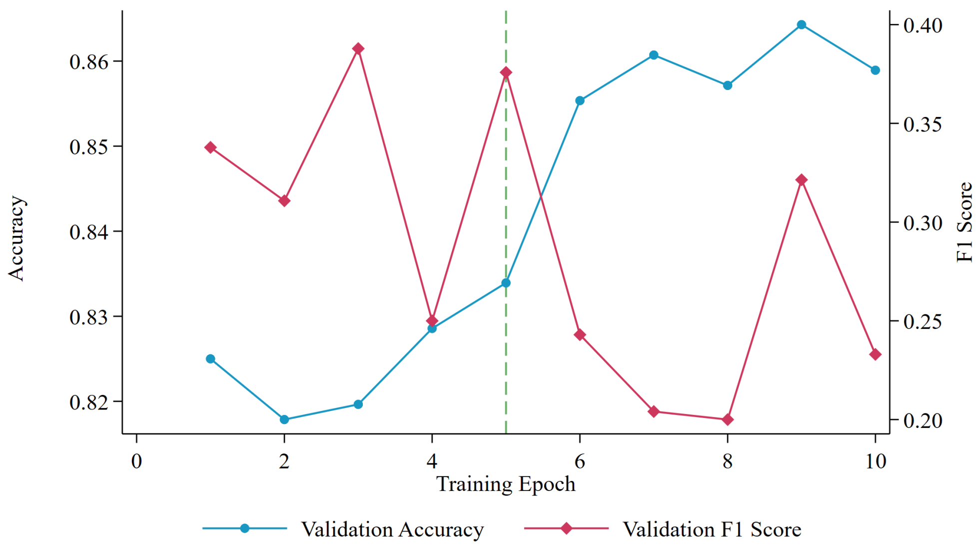
We basically face two challenges to esnure the robustness of our training process.

The first challenge was the imbalance between positive and negative samples. Since the 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 the 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, the 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 the 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 the final model, which achieved an accuracy of 0.86 and an F1 score of 0.375.

FIGURE B.3: Performance Metrics During BERT Fine-tuning for Title Attractiveness Classification



Notes: The dashed vertical line indicates the optimal checkpoint (epoch 5) selected for the final model. The model was trained on 1,900 manually annotated economics paper titles with 20% held out for validation. Training metrics were tracked using Weights & Biases system. Model performance after epoch 5 shows signs of overfitting, hence the selection of this checkpoint for subsequent analyses.

C.1. ADDITIONAL CONTROLS

To further refine the analysis and address potential confounding factors, we introduce two additional controls: article novelty and title length. Table C.1 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 C.1 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

TABLE C.1: Robustness Tests with Additional Paper Characteristics

	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	1.264***	-0.000	1.274***
	(0.001)	(0.161)	(0.001)	(0.151)
Team Citations	0.032***	3.886***	0.032***	3.876***
	(0.001)	(0.226)	(0.001)	(0.211)
Theoretical	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 Dummies	Yes	Yes	Yes	Yes
Journal FE	No	Yes	No	Yes
Affiliation Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	271,627	272,492	300,006	300,995

Notes: Novelty Index measures keyword combination atypicality following (Bramoullé and Ductor, 2018), scaled from 0 to 1. Title Length counts the number of characters in the title. Columns (1) and (3) use journal quality (ABS rating) as the dependent variable; columns (2) and (4) use citations. Corr. Citations and Team Citations are log-transformed and multiplied by 1000 for coefficient scaling. Eye Catching is a binary indicator based on BERT model classification. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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)'s 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 the 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)'s findings, suggesting that shorter titles may be more effective at capturing limited attention. However, the persistence of the title attractiveness effect indicates that attention-capturing mechanisms extend beyond simple brevity to include more sophisticated framing elements.

D.1. USING BERT LOGITS

TABLE D.1: Alternative Specifications Using Bert Logits

	Journal Quality					Citation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
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)
Eye Catching Logits	0.152***					2.283***				
	(0.006)					(0.818)				
Correspond Citations	-0.000	-0.000	-0.000	-0.000	-0.000	1.276***	1.276***	1.276***	1.276***	1.276***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.151)	(0.151)	(0.151)	(0.151)	(0.151)
Team Citations	0.032***	0.032***	0.032***	0.032***	0.032***	3.877***	3.877***	3.877***	3.877***	3.877***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.211)	(0.211)	(0.211)	(0.211)	(0.211)
Theoretical	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 Dummies	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	300,006	300,006	300,006	300,006	300,006	300,995	300,995	300,995	300,995	300,995

Notes: Eye Catching Logits uses raw prediction scores from BERT model (range 0-1) instead of binary classification. In columns (2)-(5) and (7)-(10), binary Eye Catching indicators are created using different probability thresholds (0.6, 0.7, 0.8, and 0.9 respectively). Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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

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 attention-capturing capacity.

Table D.1 presents the results of these analyses. In Columns (1) and (6), we replace the binary "Eye Catching" variable with the continuous "Eye Catching Logits" measure. The results align with the 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 (7)-(10) present the results of the threshold analysis. For journal quality (Columns 2-5), an interesting pattern emerges: the coefficient on the eye-catching measure increases as the threshold rises, from 0.130 at the 0.6 threshold to 0.134 at the 0.9 threshold. These results show titles that more effectively capture attention are associated with publication in higher-quality journals. However, the relatively small differences between coefficients suggest that the attention-capturing mechanism operates fairly consistently across different levels of title attractiveness.

For citations in Columns (7) - (10), the pattern becomes more pronounced. The coefficient increases from 2.110 at the 0.6 threshold to 2.754 at the 0.9 threshold. This substantial increase suggests that the framing effect of highly attractive titles (those clearing the 0.9 threshold) generates disproportionate benefits in terms of scholarly attention and subsequent citations.

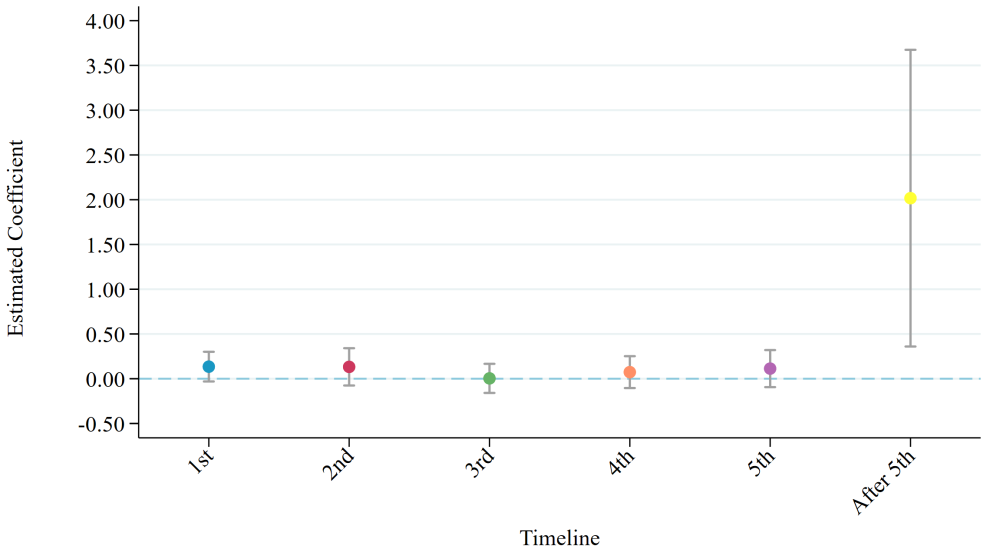
These findings offer several important insights. First, they confirm the robustness of the main results: the relationship between title attractiveness and academic outcomes persists across different measurement approaches. Second, they reveal a potential non-linear relationship, particularly for citations, where the most attention-capturing titles yield disproportionate benefits.

The differing patterns in coefficient increases between journal quality and citations illuminate how attention mechanisms function at different stages of academic impact. For journal quality, the relatively small increases suggest that while attention-capturing titles aid in publication success, this effect operates within the bounded rationality constraints of the peer review process. Once a title successfully captures initial attention, other factors like research quality become paramount.

In contrast, the citation process appears more sensitive to varying degrees of title attractiveness, reflecting different cognitive constraints and decision-making processes. Researchers facing limited attention spans often rely on titles as quick indicators of relevance when navigating vast literature. In this context, exceptionally attention-capturing titles can create powerful framing effects that significantly influence whether a paper enters a researcher's consideration set. This cognitive mechanism helps explain the more pronounced increase in citation effects at higher attractiveness thresholds, where particularly compelling titles can overcome attention barriers and establish stronger frames for their research contributions.

E.1. HETEROGENEITY OF CITATION PATTERN ON DIFFERENT YEARS

FIGURE E.1: Effect of Title Attractiveness on Citation Flows Over Time



Notes: Points show estimated coefficients for the effect of title attractiveness on citations received in each year after publication. Citations are treated as flows rather than cumulative counts. Vertical bars represent 95% confidence intervals. Estimates are from separate regressions controlling for author, journal, and paper characteristics, with journal fixed effects. The first five years are shown separately, followed by the aggregated effect for all subsequent years ("After 5th").

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 E.1 presents the estimated impact of title attractiveness on citation rates over time.

The 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 early-stage impact implies that attractive titles help papers capture initial attention in a crowded academic landscape where researchers face cognitive constraints in processing new publications. The modest but significant first-year effect suggests that attention-capturing titles can create an early advantage even before papers accumulate substantial citations.

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.

F.1. POISSON ESTIMATION

TABLE F.1: Table Appendix 1: Poisson Estimation

	Poisson							Intensive Margin	Extensive Margin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eye Catching	0.211***	0.271***	0.131***	0.141***	0.075***	0.052***	0.023	0.041***	0.008***
	(0.019)	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.007)	(0.002)
Correspond Citations			0.017***	0.011***	0.010***	0.009***		0.011***	0.000
			(0.002)	(0.002)	(0.002)	(0.002)		(0.001)	(0.000)
Team Citations			0.052***	0.057***	0.042***	0.040***	0.032***	0.036***	0.001***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
Theoretical			-0.089***	-0.123***					
			(0.020)	(0.020)					
General			0.167***	0.191***					
			(0.013)	(0.013)					
Year Dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Institution FE	No	No	No	No	No	Yes	No	No	No
Author FE	No	No	No	No	No	No	Yes	No	No
Affiliation Controls	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	325,203	325,203	301,121	285,803	286,789	243,123	240,017	259,878	286,792

Notes: Columns (1)-(7) use Poisson regression for citation count analysis. Column (8) shows intensive margin effects (citation count conditional on having at least one citation), while column (9) shows extensive margin effects (probability of receiving any citation). Corr. Citations and Team Citations are multiplied by 1000 for coefficient scaling. Eye Catching is a binary indicator based on BERT model classification. Controls include open access status, reference count, paper length, number of authors, and number of female authors. Affiliation Controls include dummy variables for institution rankings. Journal FE indicates journal-fixed effects, while Year Dummies refers to the inclusion of dummies accounting for the year of publication. Institution FE and Author FE indicate institution and author fixed effects, respectively. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

While our baseline results suggest that papers with eye-catching titles receive on average 1.925 more citations, this linear estimation may not fully capture the impact of title attractiveness on citation counts. The distribution of citations in our sample exhibits substantial right-skewness and contains numerous zero values, making the average marginal effect potentially misleading. Traditional approaches to address this issue often employ log transformations. However, recent studies (Chen and Roth, 2024; Cohn et al., 2022) demonstrate that common transformations such as $ln(x+1)$ or $arcsinh(x)$ may introduce bias in coefficient estimates and even alter their directions. To obtain more interpretable percentage effects while addressing these distributional concerns, we employ Poisson regression models:

$$E[Citations_i|X] = \exp\left(\alpha + \beta_1 \text{Attractiveness}_i + \beta_2 \text{CorrAuthor}_i + \beta_3 \text{Team}_i + \beta_4 \text{Affil}_i + X_i + \gamma_j + \lambda_t + \epsilon_i\right) \tag{F.1}$$

where all variables are defined as in Equation (1). The exponential functional form ensures non-negative predicted values while providing coefficients that can be interpreted as semi-elasticities.

Table F.1 presents our Poisson regression results. The baseline specification in Column (1) indicates that papers with eye-catching titles receive approximately 21.1% more citations. This effect remains robust and slightly increases to 27.1% after incorporating year fixed effects in Column (2). The magnitude of the effect decreases but remains statistically significant at the 1% level after progressively adding control variables in Columns (3) - (5) and journal fixed effects, consistent with our baseline findings in Table 1. Notably, Column (5) suggests that within the same journal, papers with eye-catching titles receive 7.5% more citations than those without.

The inclusion of institution fixed effects and corresponding author fixed effects in Columns (6) and (7) yields results consistent with our previous findings in Tables 3 and 4. While the coefficient becomes statistically insignificant in Column (7), this likely reflects multicollinearity from the inclusion of approximately 111,000 author dumour variables rather than a genuine absence of effect. This interpretation is particularly reasonable given that citation decisions typically do not heavily weight author identity, as discussed in our main analysis.

Following Chen and Roth (2024), we further decompose the effect into intensive and extensive margins to better understand the mechanisms through which title attractiveness affects citations. Our intensive margin analysis in Column (8), which excludes zero-citation papers and employs a log transformation of citations, reveals that among cited papers, eye-catching titles are associated with a 4% increase in citation count. The extensive margin analysis in Column (9), which transforms citations into a binary outcome indicating whether a paper is cited, shows that papers with eye-catching titles are 0.8 percentage points more likely to be cited.