What are economists telling us in the top 5 ? A Structural Topic Models approach

Si Hao Li
ENS Paris Saclay
si.li@ens-paris-saclay.fr

6 novembre 2024

Résumé

This article examines the evolution of research topics published in top five economics journals using advanced textual analysis methods. We implemented Structural Topic Model (Roberts et al., 2014) on our database of abstracts from the top five journals, which we constructed using web scraping techniques from the RePEc platform. Our findings indicate that economics has evolved, with an increasing resurgence of empirical methods and a significant decline in theory, particularly in general equilibrium.

Keywords : NLP, Structural Topic Modelling, economists, econometrics, top five, AER
JEL classification : A11 ; B41 ; C01

1 Introduction

Publishing in one of the top five economics journals is often perceived as the pinnacle of an economist's career, due to the difficulty of being selected and the lengthy peer-review process that can extend over several years, sometimes up to five (Okret-Manville et al., 2015). This situation underscores the considerable influence these journals have on researchers' career trajectories. Indeed, the work of Heckman and Moktan (2020) illustrates this "top five" tyranny, where economists who publish one, two, or three articles in these journals see their chances of obtaining tenure in the United States increase exponentially. Similar observations have been made in Europe, as shown by Graber and Wälde (2008), suggesting a global trend towards this "top five" race.

In this article, we consider the "top" economics journals to be composed of the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies*, in accordance with the literature on this subject (Pieters and Baumgartner, 2002; Card and DellaVigna, 2013; **?**; Heckman and Moktan, 2020).

The economics literature has extensively examined the "top five" phenomenon. Some studies criticize the disproportionate influence of these journals on the direction of economic research, suggesting a potential bias towards certain topics, methodologies, or authors at the expense of intellectual diversity (Fourcade et al., 2015). Other studies adopt a more descriptive approach, analyzing publication trends, co-citation networks, or the geographical and institutional distribution of authors (Hamermesh, 2018; Linnemer, 2024). Additionally, textual analyses have been conducted to examine the dominant themes in these journals, although often limited to specific periods or restricted samples (Wei, 2018).

However, several gaps remain in the literature, particularly when it comes to combining the last two research areas. First, most studies focus on specific time periods, preventing a comprehensive and evolving understanding of long-term research trends. Second, existing thematic analyses do not fully leverage modern text analysis methods, notably in machine learning, such as *Topic Modeling* and its extensions, which can integrate metadata and capture complex dynamics.

The objective of this article is to exhaustively analyze the evolution of research themes published in the five most prestigious economics journals over an extended period. To achieve this, we use *Structural Topic Modeling* on the abstracts of articles—a methodology that allows for the discovery of latent themes while incorporating contextual information or metadata (Roberts et al., 2014). Moreover, it is particularly suited for processing large text corpora, which is essential given the considerable volume of articles published in the "top five" over the decades.

Our study contributes to the literature in several ways. First, by covering an extended time period, we provide an in-depth historical perspective on the evolution of research topics in economics. Second, by applying STM, we demonstrate the utility of advanced text analysis methods for exploring scientific corpora, paving the way for more sophisticated analyses. Third, by shedding light on the internal thematic dynamics of the "top five," we offer insights into how the orientations of economic research are shaped within the most influential journals themselves. Thus, our study adopts an interdisciplinary approach, combining economics, scientometrics, and text analysis, to enrich our understanding of scientific production in economics.

2 Method

2.1 Data

Our database was constructed using web scraping techniques applied to the RePEc platform (Research Papers in Economics), an initiative aimed at cataloging and optimally providing access to economic research works. We focused on analyzing the five top economics journals as mentioned in the introduction : the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of*

Economic Studies. We collected these research articles by specifically concentrating on their titles and abstracts. Each article is represented by its main metadata, including the title, abstract, authors, citation, year of publication, and journal.

In total, the dataset contains 13,271 articles. However, some journals are not indexed before the 1980s or lack complete abstracts. Consequently, we chose to limit our analysis to the period from 1980 to 2023, relying on a corpus of 11,140 articles. In this corpus, the average text length is 115.5 words with a standard deviation of 39.53, indicating notable dispersion. The most concise texts contain 15 words, while the longest can reach up to 781 words. Table 1 presents the descriptive textual statistics for the various journals and periods. It shows a relatively similar number of articles, except for the AER, which has approximately 1,000 more articles than the other journals. Additionally, the average number of words per abstract ranges between 104 and 125, depending on the publication.

Table 1 – Textual descriptive statistics of database

Variable	**Source**	**Mean**	**SD**	**Min**	**Max**	**Article Count**
Journal						
1	AER	104.55	31.80	19	354	3389
2	ECON	124.18	45.09	19	394	2082
3	JPE	122.02	41.44	18	325	1821
4	QJE	107.22	27.26	21	329	1757
5	RES	125.88	45.55	15	781	2091
Period						
1980-1990	Period1	103.35	38.80	15	781	1343
1991-2000	Period2	105.61	32.46	19	320	1910
2001-2010	Period3	123.22	42.14	26	394	2702
2011-2020	Period4	117.61	39.10	22	354	3890
2020-2023	Period5	124.04	37.64	30	328	1212

2.2 Model

Natural language processing methods are increasingly used in social sciences, particularly in economics (Gentzkow et al., 2019). We have identified several types of applications in economic research, such as forecasting economic activity (Aprigliano et al., 2023), predicting financial markets (Xing et al., 2018), as well as other uses in political economy, notably through sentiment analysis (Algan and Renault, 2024). In this article, we continue this trend of using text as data by employing a topic model—a probabilistic model that allows for the identification of latent subjects or themes within a text corpus. The reference model often used is Latent Dirichlet Allocation (LDA) (Blei et al., 2003), whose algorithm iteratively calculates the probabilities of tokens belonging to topics and topics to documents, enabling us to visualize the composition of the identified subjects. In this LDA model, parameter estimation relies on the posterior inference of latent variables from observed data. This approach can be interpreted as the maximization of the logarithmic likelihood, implemented through an Expectation-Maximization algorithm, typically used in mixture models. In this article, we will use Structural Topic Modeling (Roberts et al., 2014), which, in addition to detecting underlying themes in a text corpus like LDA, allows for the integration of additional information, called covariates.

2.2.1 Choise of K

Thus, we employ Structural Topic Modeling to analyze our article abstracts from the top five economics journals. STM allows us to identify latent topics within the documents while taking into account metadata such as the journal and the year of publication. Initially, we preprocess the data by cleaning it and tokenizing the terms. We then test several values of K to determine the optimal number of topics and calculate four distinct metrics : the held-out likelihood, which measures the probability of occurrence of words removed during estimation (to be maximized) ; residuals, which assess the over-dispersion of the residuals (to be minimized) ; semantic coherence, which favors the co-occurrence of words within the same topic (to be maximized) ; and finally, the lower bound, which allows us to monitor the convergence of the model (to be minimized).

We adopt an approach that prioritizes the semantic coherence of topics rather than focu-

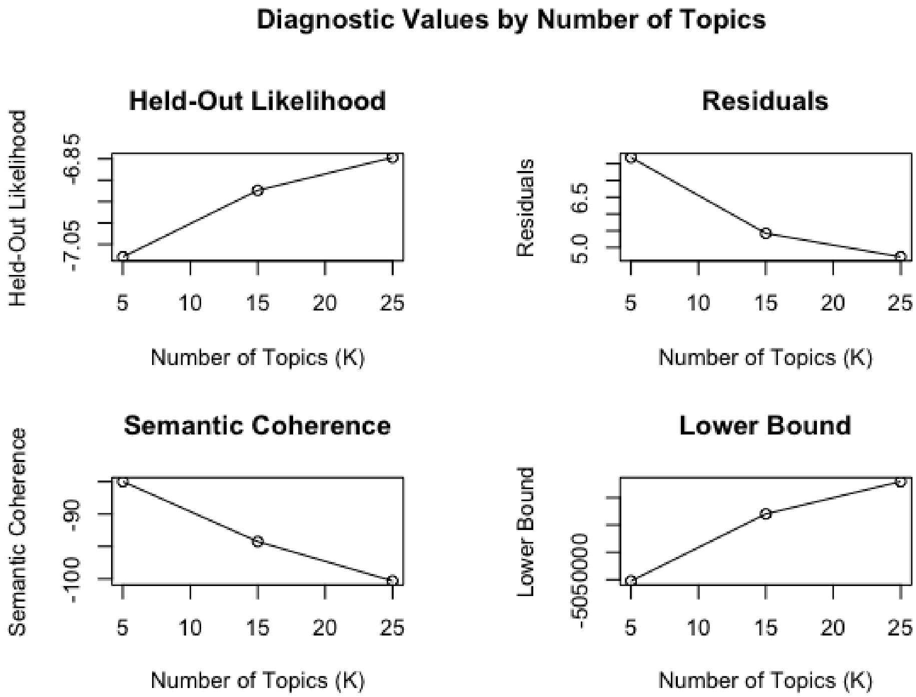

FIGURE 1 – Metrics for NoT

sing exclusively on optimized model performance. Indeed, in the context of specific themes in economic sub-disciplines such as econometrics, development economics, or industrial organization, as well as other related fields, it would be inappropriate to use a high value of the parameter K for metric calculation. Such an approach would risk compromising the quality of topic interpretation by introducing excessive fragmentation.

2.2.2 Structural Topic Modeling

The STM models the generation of words in documents through a two-step probabilistic process. First, for each document d, the topic proportions $\theta_d = (\theta_{d1}, \theta_{d2}, \ldots, \theta_{dK})$ are drawn from a *logistic-normal* distribution conditioned on the covariates associated with the document. Second, for each word n in document d, a topic z_{dn} is selected according to a multinomial distribution with proportions θ_d, and then the word w_{dn} is generated from the word distribution associated with the topic z_{dn}, denoted $\phi_{z_{dn}}$.

The topic proportions for each document d are modeled using a logistic-normal distribution, which allows for the incorporation of covariates and modeling correlations between topics. The logit transformation ensures that the topic proportions are between 0 and 1 and that their sum equals 1.

Logit We define the latent vector $\boldsymbol{\eta}_d$ by :

$$\eta_{dk} = \log\left(\frac{\theta_{dk}}{\theta_{dK}}\right), \quad \text{for } k = 1, \ldots, K-1 \tag{1}$$

where θ_{dK} is the proportion of the reference topic K.

Linear model with Covariates The vector $\boldsymbol{\eta}_d$ is then modeled as a function of the covariates according to the following equation :

$$\eta_{dk} = \beta_{0k} + \boldsymbol{\beta}_k' \mathbf{x}_d + f_k(\text{Year}d) + \epsilon dk \tag{2}$$

In this equation, β_{0k} represents the intercept for topic k. The vector $\boldsymbol{\beta}_k$ corresponds to the coefficients associated with the covariates $\mathbf{x}_d$, which may include, for example, indicator variables for the journal. The function $f_k(\text{Year}d)$ is a spline capturing the nonlinear effect of the year on topic k. Finally, the error term ϵdk is assumed to follow a normal distribution $\mathcal{N}(0, \sigma^2)$.

Multinomiale For each document d, words are generated according to a multinomial distribution conditioned on the proportions of topics θ_d and the word distributions of topics ϕ_k :

$$\mathbf{w}_d \sim \text{Multinomiale}\left(N_d, \sum_{k=1}^{K} \theta_{dk}\phi_k\right) \tag{3}$$

In this equation (3), $\mathbf{w}d$ is the vector of word counts in document d, and N_d represents the total number of words in this document. The ϕ_k distribution corresponds to the word distributions for topic k, with $\phi kw = P(w \mid z = k)$."

Likelihood The estimation of the model parameters (β_{0k}, $\boldsymbol{\beta}_k$, $f_k(\cdot)$, ϕ_k) is performed by maximizing the joint likelihood of the observed data :

$$\mathcal{L} = \prod_{d=1}^{D} \left[\int_{\theta_d} \left(\prod_{n=1}^{N_d} \sum_{k=1}^{K} \theta_{dk} \phi_{kw_{dn}} \right) p(\theta_d \mid \mathbf{x}_d), d\theta_d \right] \tag{4}$$

where $p(\theta_d \mid \mathbf{x}_d)$ is the density of the logistic-normal distribution of the topic proportions.

3 Results

3.1 Topic analysis

As mentioned in the introduction, we tested several values of K to determine an optimal number of topics for our analysis. Based on the results presented in Figure 1, we opted for a value of $K = 15$. Although this choice is not the most efficient from a strictly optimization standpoint, it allows us to capture a coherent and interpretable set of themes. In reality, there is no definitive answer regarding the ideal number of topics to retain (Grimmer and Stewart, 2013). The initial calculation provides a preliminary estimate, but it is essential to proceed with a manual evaluation by examining each searchK test we performed.

Table 2 presents the different topics, accompanied by four metrics used to identify the most representative words of each topic. The first metric, "Highest Probability," evaluates the highest probability of a word appearing in a given topic. "FREX" measures the exclusivity of a word within each topic. "Lift" highlights words that are globally rare but relatively frequent in the specific topic. Finally, the "Score" serves as a synthesis of these different metrics.

This approach enables us to identify certain topics that are apparent, such as Topic 11, which pertains to financial economics, and Topic 1, related to macroeconomics. For the less evident topics, we search for the most representative abstracts for each topic based on the probabilities of documents belonging to each topic. This process allows us to define each topic presented in Table 3 by illustrating it with three reference articles specific to each topic.

For example, Topic 9 aptly demonstrates this methodology : by conducting an in-depth qualitative analysis of the associated articles, we primarily identified studies focusing on Nash equilibrium, games with incomplete information, and auction theory, among others. Consequently, we deduced that Topic 9 is associated with game theory.

To more precisely evaluate the relevance of the studied themes, Figure 2 presents the 15 subjects ranked according to their prevalence—that is, the proportion of documents dedicated to each. We observe that the topics of econometrics and international economics occupy the largest share, followed by applied microeconomics (Topic 5) and macroeconomics. Although it is challenging to assert the existence of an "empirical turn" (Angrist et al., 2017), the significant emphasis placed on econometrics and empirically oriented related disciplines suggests that economic science is no longer confined to a theoretical approach. It now tends to confront its theories with empirical evidence to test their robustness.

Figure 3 illustrates the evolution of the expected proportions of the fifteen topics in top economic journals. Several significant trends emerge. On one hand, themes related to health economics, labor and education economics, as well as development economics and geographical, institutional, and historical approaches show a constant progression. This reflects an increased interest in the empirical and societal dimensions of economic research. This empirical orientation contributes to a growing hyperspecialization among researchers (Périvier, 2020), potentially reducing economists' ability to maintain an overarching view of their discipline (Coase, 2000) [1].

On the other hand, more theoretical themes, such as Topic 14 (including general equilibrium and mathematical economics), are decreasing in prevalence, suggesting a shift toward

1. *"Il se dégage de tout cela que les économistes se considèrent comme les détenteurs d'une boîte à outils mais non comme concernés par un objet d'étude spécifique..."* Coase, R. H. (2000). L'économie néo-institutionnelle. Revue d'économie industrielle, 92(1), 51-54.

Table 2 – Top Words for Each Topic

Topic	Metric	Top Words
Topic 1	Highest Prob	model, rate, shock, price, aggreg, polici, chang
	FREX	inflat, fluctuat, nomin, shock, cyclic, monetari, unemploy
	Lift	-digit, backus, backus-smith, bil, bookstor, bretton, burdett
	Score	shock, unemploy, inflat, monetari, price, rate, fluctuat
Topic 2	Highest Prob	econom, countri, local, growth, use, data, network
	FREX	migrat, land, farm, centuri, patent, farmer, agricultur
	Lift	agro-climat, arctic, berlin, bowl, coman, deeply-root, dust
	Score	citi, countri, ethnic, agricultur, network, patent, growth
Topic 3	Highest Prob	polici, polit, group, govern, vote, social, voter
	FREX	vote, voter, elect, candid, elector, politician, corrupt
	Lift	-parti, agenda-set, anti-corrupt, anticorrupt, apprehend, ask", bayli
	Score	polit, voter, vote, elect, parti, politician, elector
Topic 4	Highest Prob	version, anoth, borrow, abstract, item, itemthi, paper
	FREX	anoth, abstract, item, itemthi, lectur, this, decemb
	Lift	capitalthi, effectsthi, growththi, lectur, this, underbid, item
	Score	abstract, version, anoth, borrow, itemthi, item, alfr
Topic 5	Highest Prob	inform, agent, contract, can, incent, mechan, optim
	FREX	learn, communic, renegoti, inform, reput, princip, agent
	Lift	-hous, "learn, "strateg, better-inform, collusion-proof, concavif, contagi
	Score	agent, inform, contract, incent, signal, optim, learn
Topic 6	Highest Prob	risk, consumpt, household, return, asset, price, incom
	FREX	portfolio, consumpt, wealth, risk, stock, household, annuiti
	Lift	casino, consumption-wealth, crra, felic, giffen, hunger, jones
	Score	risk, consumpt, household, asset, wealth, stock, portfolio
Topic 7	Highest Prob	effect, use, estim, find, program, increas, health
	FREX	health, hospit, drug, medicar, treatment, medicaid, physician
	Lift	caregiv, curat, ltchs, overbil, overus, pneumonia, pre-trend
	Score	health, treatment, insur, mortal, hospit, estim, program
Topic 8	Highest Prob	wage, worker, labor, school, employ, educ, job
	FREX	women, men, gender, femal, marriag, male, school
	Lift	animus, arreste, baccalaur, birthrat, bride, buse, callback
	Score	worker, school, wage, women, student, children, job
Topic 9	Highest Prob	equilibrium, game, price, equilibria, player, auction, strategi
	FREX	auction, bid, bidder, equilibria, seller, nash, buyer
	Lift	-pay, "demand, action-sampl, alternating-mov, bertrand-edgeworth, bertrand-lik, buyer'
	Score	game, player, equilibria, auction, equilibrium, seller, buyer
Topic 10	Highest Prob	estim, model, test, use, distribut, method, paramet
	FREX	asymptot, nonparametr, bootstrap, semiparametr, null, regressor, carlo
	Lift	column, cross-valid, fiml, kullback-leibl, mcmc, near-integr, nfxp
	Score	asymptot, estim, nonparametr, method, econometr, regressor, paramet
Topic 11	Highest Prob	financi, bank, market, credit, debt, liquid, asset
	FREX	bank, credit, debt, loan, crisi, crise, mortgag
	Lift	-borrow, -mile, asset-back, atlanta, bailey, bailout, bank-level
	Score	bank, debt, credit, financi, liquid, default, borrow
Topic 12	Highest Prob	tax, state, incom, percent, increas, unit, rate
	FREX	tax, ceo, union, spend, takeov, reform, compani
	Lift	anti-tax, belt, carryov, city-pair, deced, evad, evas
	Score	tax, incom, spend, percent, reform, fiscal, state
Topic 13	Highest Prob	firm, product, trade, market, price, model, cost
	FREX	export, firm, tariff, product, retail, industri, multin
	Lift	ldcs, graviti, across-firm, apparel, arkolaki, arment, autom
	Score	firm, product, trade, industri, export, technolog, price
Topic 14	Highest Prob	copyright, paper, model, optim, univers, author, press
	FREX	press, chicago, author, univers, oxford, copyright, commod
	Lift	barro-beck, benvenist, chamley, divisia, duction, equity-effici, externality-correct
	Score	copyright, press, chicago, author, univers, optim, economi
Topic 15	Highest Prob	prefer, choic, model, behavior, decis, util, belief
	FREX	axiom, choic, prefer, belief, ambigu, subject, util
	Lift	"prefer, afriat, allai, allot, ambiguity-avers, ambiguity-sensit, anscomb
	Score	prefer, choic, belief, util, axiom, experiment, subject

Table 3 – References for each topic

Topic	Topic name	References
Topic 1	Macroeconomics	Steinsson (2008) ; Bils et al. (2012) ; Eichenbaum et Evans (1995)
Topic 2	Development Economics / Geography, Institutions, Historical approach	Alesina et al. (2016) ; Ashraf et Galor (2013) ; Black et al. (2015)
Topic 3	Political economy	Bernhardt et al. (2020) ; Acemoglu et al. (2013) ; Grossman et Helpman (1996)
Topic 4	Economist things	Hurwicz (2008) ; Mundell (1999)
Topic 5	Contract Theory / Mechanism Design / Industrial Organization	Halac et Yared (2020) ; Ely et Szydlowski (2020) ; Ben-Porath et al. (2014)
Topic 6	Financial Economics : Asset Pricing	Campbell et Vuolteenaho (2004) ; Parker (2003)
Topic 7	Health Economics	Almond et al. (2010) ; East et al. (2023) ; Finkelstein et al. (2019)
Topic 8	Labor economics / Education economics / Inequality	Bayer et Charles (2018) ; Herr (2015) ; Juhn et McCue (2016)
Topic 9	Game Theory	Dufwenberg et Stegeman (2002) ; McAdams (2003) ; Lauermann et Speit (2023)
Topic 10	Econometrics	Andrews (2016) ; Cavaliere et Georgiev (2020) ; Paparoditis et Politis (2003)
Topic 11	Financial economics : macroeconomics approach	Koch et al. (2016) ; Chang et Velasco (2001) ; Brunnermeier et al. (2016)
Topic 12	Public Economics	Goolsbee (2000) ; Kreiner et al. (2014) ; Dubin et al. (1992)
Topic 13	International Economics	Mayer et al. (2014) ; Bernard et al. (2019) ; Goldberg et al. (2010)
Topic 14	Economic theory (General Equilibrium, mathematical economics)	Grossman et Kim (1995) ; Hori (1982) ; Laitner (1990)
Topic 15	Decision theory	Gul et al. (2014) ; Hill (2019) ; Baillon et al. (2011)

more applied and empirical work. Topics like econometrics, macroeconomics, and asset pricing are also showing a decline. For econometrics, this trend could be attributed to a reduction in theoretical publications on econometric methods, as confirmed in Garg and Fetzer (2024), or to a gradual normalization of their use, where researchers, although applying econometric techniques, omit to mention them explicitly in the abstracts.

3.2 Review analysis

In this section, we analyze the temporal evolution of thematic proportions within the top five journals. This can be interpreted as a reflection of increasing internal diversification within each main topic.

Consider a main topic T_k that is decomposed into several sub-topics $T_{k,i}$, where i indexes the sub-topics of topic k. The observed proportion for topic T_k in a journal at a given date can be modeled by the sum of the proportions of its individual sub-topics :

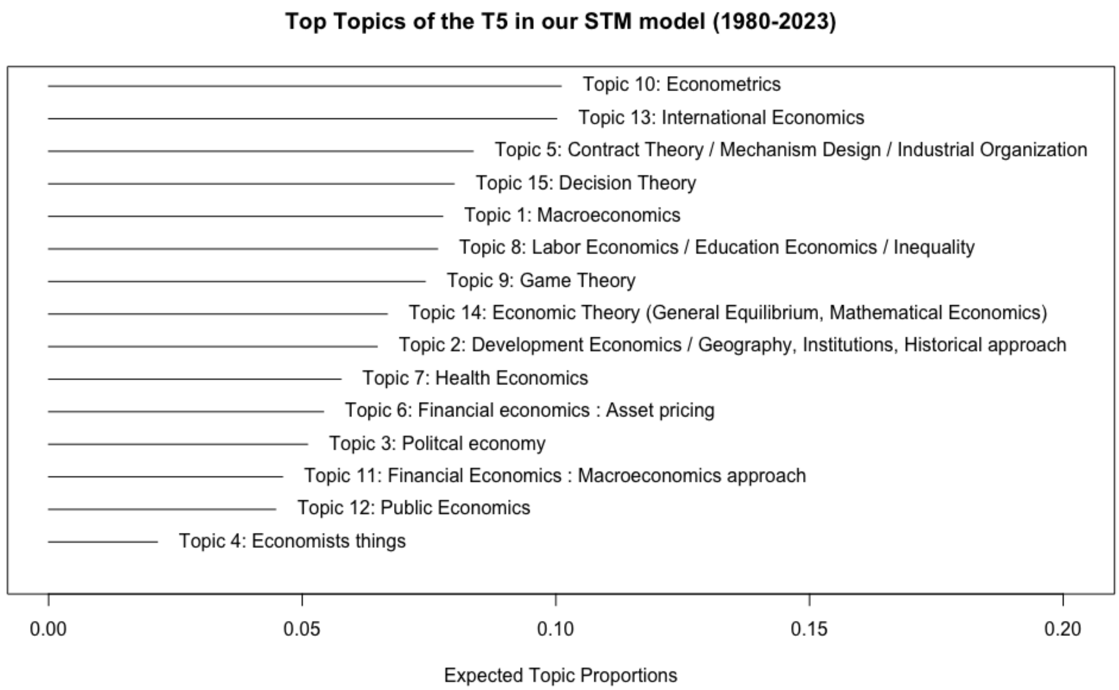

FIGURE 2 – Top Topics of the T5 in out STM

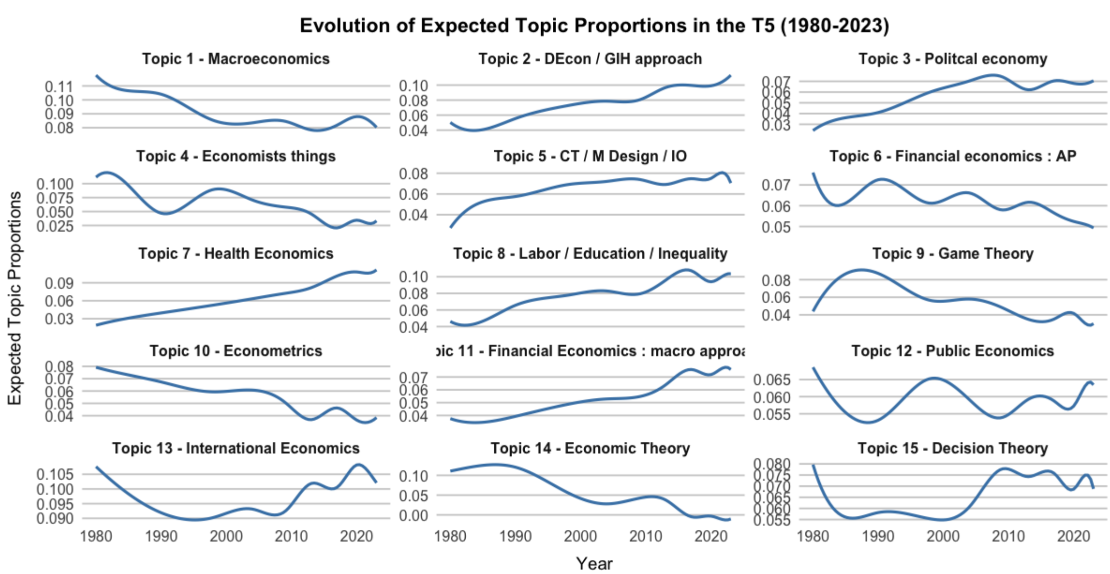

FIGURE 3 – Evolution of ETP in the T5 (1980-2023)

$$P(T_k) = \sum_i P(T_{k,i}),$$

where $P(T_{k,i})$ represents the proportion of sub-topic $T_{k,i}$ within the main topic T_k. In this context, if the number of sub-topics i increases over time, each proportion $P(T_{k,i})$ may decrease even if the overall interest in topic T_k remains stable or increases. This reflects a dilution of the total proportion $P(T_k)$, due to its distribution over a greater number of sub-topics, rather than a decrease in the importance of the main topic.

Thus, Figure 4 presents the evolution of the expected proportions of topics, as analyzed by our Structural Topic Model applied to the abstracts of published articles. The analysis highlights a general decrease in the proportion of dominant topics between 1980 and the early 2000s, followed by stabilization or even a slight increase over the past two decades. We observe that the American Economic Review maintains a higher proportion of topics than the other journals, which seems to indicate a more generalist and diversified approach. The Quarterly Journal of Economics and the Journal of Political Economy display intermediate proportions, while Econometrica and the Review of Economic Studies present relatively lower proportions, probably due to their more specific focus on specialized methodologies and theories.

Figure 5 presents the estimated proportions for each theme, accompanied by their corresponding confidence intervals, which allows for a statistical evaluation of the differences in thematic emphasis between the AER and the other journals. The thematic differences observed in Figure 2 highlight distinct thematic profiles between the AER and the other journals. Although some themes are common to both the AER and the other journals, each journal exhibits specific thematic characteristics. For example, Themes 5, 6, and 7 appear with increased prevalence in ECON compared to the AER. The QJE seems to have a thematic distribution relatively similar to that of the AER.

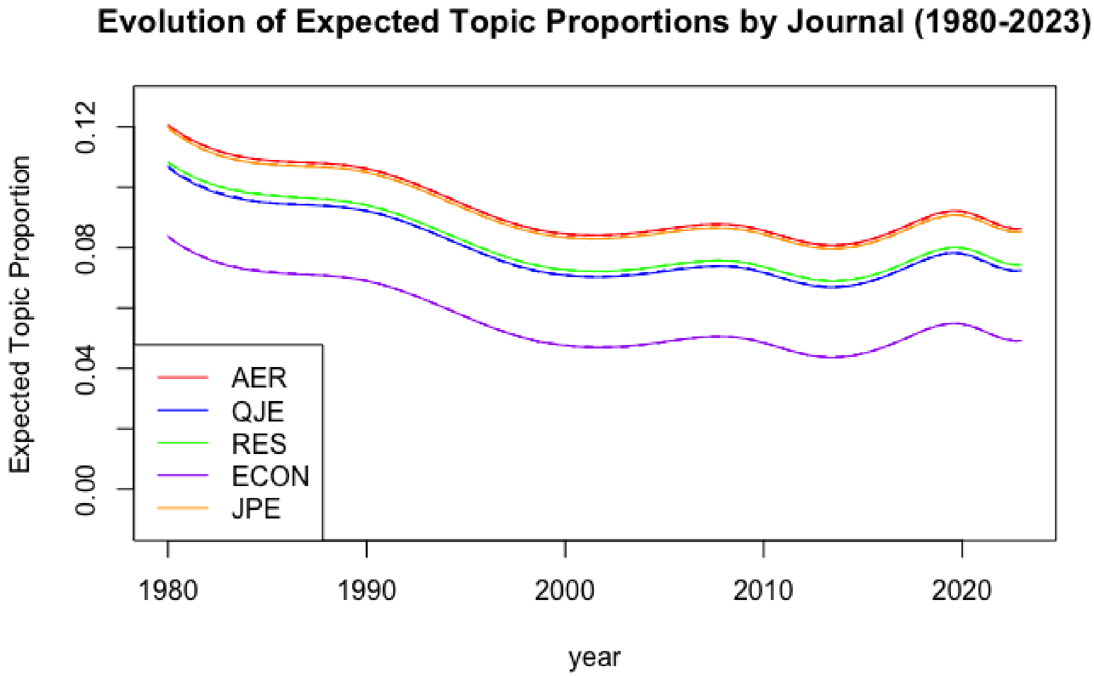

FIGURE 4 – Evolution of ETP by Journal (1980-2023)

3.3 Topic Correlation

To deepen the analysis of the relationships between the different identified topics, we present a topic correlation graph and a correlation matrix. These tools allow us to visualize thematic links and quantify the correlations between subjects, thereby offering a better understanding of the overall structure of the research domains.

Figure 6 illustrates the topic correlation graph, where each node represents a topic and the links between nodes indicate significant correlations. The observed clusters confirm coherent thematic groupings. For example, the topics related to macroeconomics (Topic 1) and financial economics with a macroeconomic approach (Topic 11) show strong connectivity, reflecting a natural interdependence between these two domains. Similarly, Topics 14, 9, 6, and 15, which pertain to theoretical or applied microeconomics, form a distinct cluster. Another node emerges around Topic 12, representing sub-disciplines that have emerged following the empirical turn we have discussed.

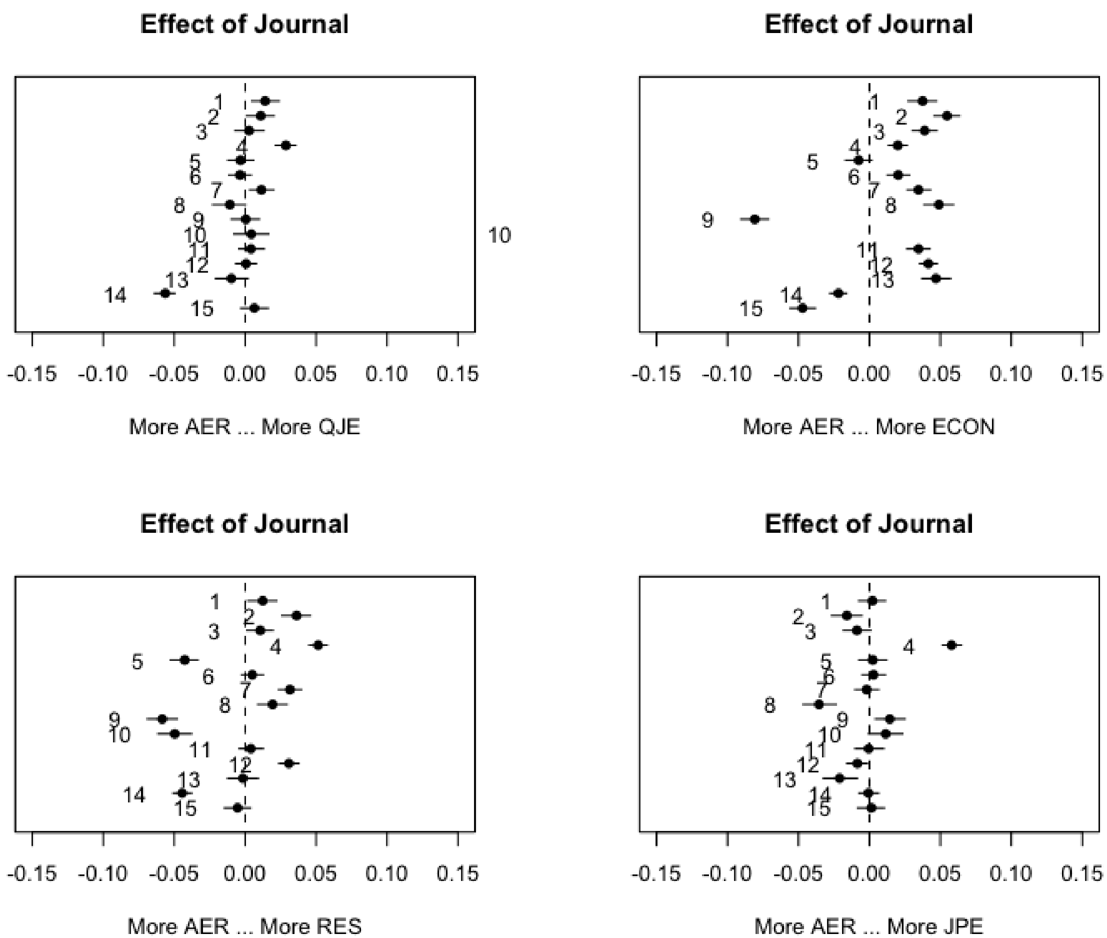

Figure 5 – AER vs.

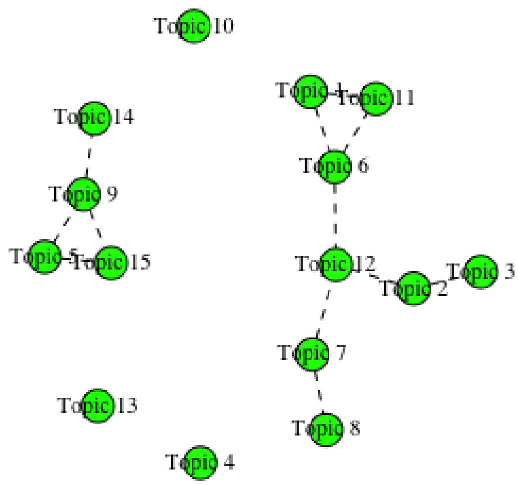

FIGURE 6 – Topic correlations

Figure 7, which presents the topic correlation matrix, provides a quantitative perspective of the pairwise relationships between topics, where each color indicates the intensity of the correlation. High values (light red areas off the diagonal) highlight strong thematic overlaps, such as between macroeconomics and finance, or between "Public Economics" (Topic 12) and development economics. Lower correlations (blue or purple areas) indicate less related subjects, such as “Health Economics” (Topic 7) and “Game Theory” (Topic 9), which appear independent.

The model’s ability to capture logical relationships between connected research fields, while clearly differentiating distinct subdomains, demonstrates its robustness in determining our topics.

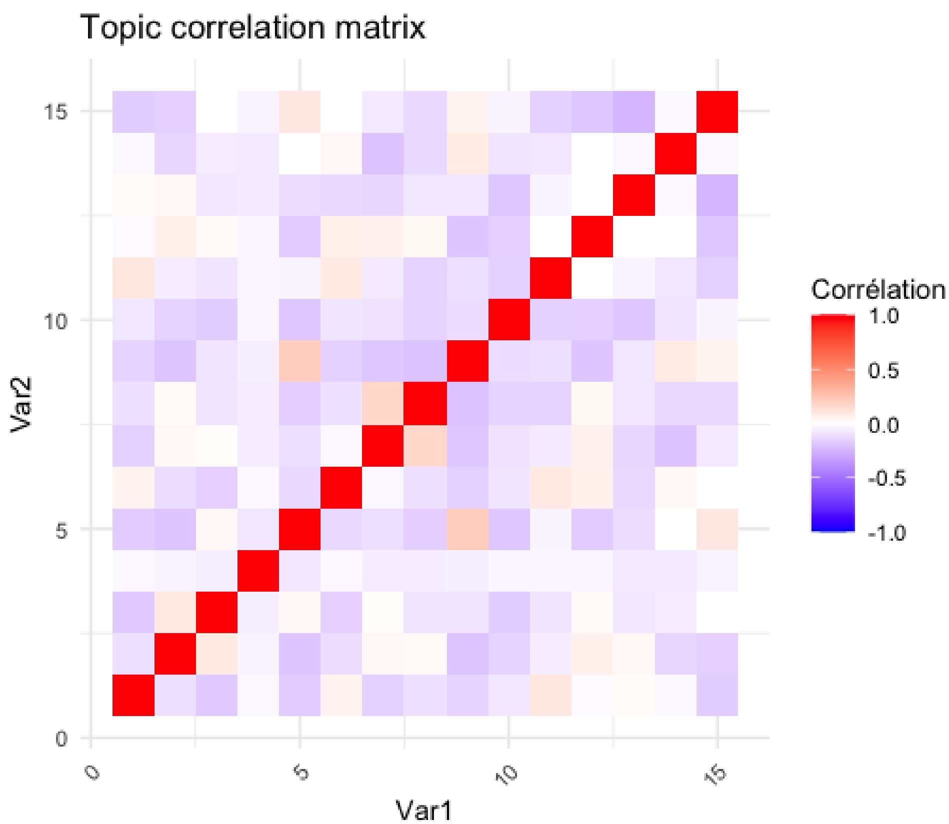

Figure 7 – Topic correlation matrix

4 Conclusion

This article employs various data analysis tools, such as web scraping and machine learning, to investigate the evolution of topics in economics since 1980. The use of the Structural Topic Model allowed for an in-depth classification of subjects, overcoming challenges linked to gaps in the longitudinal data of JEL codes. Although alternative computational social science methods could have been used—such as employing large language models (LLMs) through fine-tuning BERT (Do et al., 2024) or zero-shot classification (Gilardi et al., 2023)—the strong performance of our STM justified prioritizing this approach.

The contribution of this article lies, on one hand, in the construction of a structured database and, on the other, in the illustration of computational methods that may be underutilized in economics. Our model's findings highlight the most prominent themes since 1980 and qualitatively depict the evolution of publications within the field's most prestigious journals. Echoing several prior studies, our results confirm the empirical turn observed within the discipline.

We also examined thematic diversity within the "Top Five" economic journals, which appears to have expanded since 1980, with a recent stabilization. This increased diversity could, in our view, stem directly from this empirical shift.

Looking forward, our method could be applied in the fields of the economics of science and gender studies to analyze female economists' publishing practices : are they publishing in the "Top Five" ? If so, in which domains ? Such analyses could open valuable perspectives on gender dynamics within academic publishing in economics.

Références

Algan, Y. and Renault, T. (2024). La france sous nos tweets—portrait d'une france en colère, et de ses conséquences politiques. Technical report, CEPREMAP.

Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2017). Economic research evolves : Fields and styles. *American Economic Review*, 107(5) :293–297.

Aprigliano, V., Emiliozzi, S., Guaitoli, G., Luciani, A., Marcucci, J., and Monteforte, L. (2023). The power of text-based indicators in forecasting italian economic activity. *International Journal of Forecasting*, 39(2) :791–808.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan) :993–1022.

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

Do, S., Ollion, É., and Shen, R. (2024). The augmented social scientist : Using sequential transfer learning to annotate millions of texts with human-level accuracy. *Sociological Methods & Research*, 53(3) :1167–1200.

Fourcade, M., Ollion, E., and Algan, Y. (2015). The superiority of economists. *Journal of Economic Perspectives*, 29(1) :89–114.

Garg, P. and Fetzer, T. (2024). Causal claims in economics. Working Paper.

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

Gilardi, F., Alizadeh, M., and Kubli, M. (2023). Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30) :e2305016120.

Graber, M. and Wälde, K. (2008). Publish or perish ? the increasing importance of publications for prospective economics professors in austria, germany and switzerland. *German Economic Review*, 9(4) :457–472.

Grimmer, J. and Stewart, B. M. (2013). Text as data : The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3) :267–297.

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

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.

Linnemer, L. (2024). A menagerie of rankings : a look in repec's factory. *Scientometrics*, 129(1) :321–372.

Okret-Manville, C., Chapoy, E., and Fraysse, M. (2015). Les working papers en économie. *I2D-Information, données & documents*, 52(1) :52–53.

Périvier, H. (2020). *L'économie féministe*. Presses de Sciences Po.

Pieters, R. and Baumgartner, H. (2002). Who talks to whom ? intra-and interdisciplinary communication of economics journals. *Journal of Economic Literature*, 40(2) :483–509.

Roberts, M. E., Stewart, B. M., Tingley, D., et al. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4) :1064–1082.

Wei, H. (2018). Understanding the dynamics of top-tier economics publications. *Economics Letters*, 163 :45–48.

Xing, F. Z., Cambria, E., and Welsch, R. E. (2018). Natural language based financial forecasting : a survey. *Artificial Intelligence Review*, 50(1) :49–73.

A Appendix : Metrics

This appendix provides a detailed explanation of the various metrics used to evaluate words within the identified topics in our analysis.

A.1 Exclusivity (FREX)

The FREX metric (Bischof & Airoldi, 2012) summarizes a word's contribution to a topic by using the harmonic mean of its rank in terms of exclusivity and frequency. This approach prevents a high rank in one dimension from compensating for a low rank in the other. FREX is defined as :

$$\text{FREX} = \left(\frac{w}{F} + \frac{1-w}{E} \right)^{-1}$$

where : - F is the rank of the word in terms of frequency within the topic, - E is the rank of the word in terms of exclusivity to the topic, - w is a weighting parameter between 0 and 1, balancing the importance of frequency and exclusivity.

A.2 Lift

The Lift metric highlights words that are globally rare but relatively frequent in a specific topic. Lift is defined as :

$$\text{Lift}(w_i, t_j) = \frac{P(w_i|t_j)}{P(w_i)}$$

where $P(w_i)$ is the probability of word w_i across the entire corpus, and $P(w_i|t_j)$ is the probability of word w_i within topic t_j. A high lift score indicates that the word is more specific to topic t_j.

B Data annotation and Gender

In this section, we present a methodology for annotating a textual corpus using the Mistral AI API. This process aims to determine and categorize the presence or absence of female authors in the database. We apply the prompt method described in the appendix of the article by F. Gilardi et al. (2023), as it offers a systematic and reproducible approach to text annotation.

B.1 Code Example

```python
import os
import csv
import time
from tqdm import tqdm

from mistralai.client import MistralClient
from mistralai import Mistral, UserMessage

model = "mistral-small-latest"

client = Mistral(api_key="KEY")

prompt = """For this task, I am asking you to annotate the authors. For
    each group of authors, follow these instructions:
Read the group of authors carefully, paying close attention to details.
    Avoid using shortcuts.
Sort the list of authors and try to determine if there is a woman in the
    group or not. You should simply respond with one of the two terms:
    Woman or No Woman.
Authors should be annotated as Woman when the list contains at least one
    woman.
```

```
Authors should be annotated as No Woman when the list contains no women.
    """

def annotate_sentence(sentence):
    full_prompt = prompt + sentence
    chat_response = client.chat.complete(
        model=model,
        messages=[
            {
                "role": "user",
                "content": full_prompt,
            },
        ]
    )
    return chat_response.choices[0].message.content

input_csv = "data_top5_without.csv"
output_csv = "new_df_top5.csv"

with open(input_csv, 'r', encoding='utf-8') as f:
    total_rows = sum(1 for line in f) - 1

with open(input_csv, mode='r', newline='', encoding='utf-8') as infile, \
     open(output_csv, mode='w', newline='', encoding='utf-8') as outfile:

    reader = csv.DictReader(infile)
    fieldnames = reader.fieldnames + ['Annotation']
    writer = csv.DictWriter(outfile, fieldnames=fieldnames)

    writer.writeheader()

    for row in tqdm(reader, total=total_rows):
        sentence = row['authors']
```

```
        annotation = annotate_sentence(sentence)
        row['Annotation'] = annotation
        writer.writerow(row)
        time.sleep(0.75)

print(f"Les phrases annot es ont  t  sauvegard es dans {output_csv}.")
```

B.2 Evaluating annotations

To assess the effectiveness and quality of the annotation, it is common practice to use scores, particularly the F1 score and either Fleiss' Kappa (for more than 2 annotators) or Cohen's Kappa (for 2 annotators), weighted by the annotation error rate. This rate can vary significantly depending on the prompt used. The F1 score is a measure of the prediction accuracy of a classification model. It is the harmonic mean of precision, which measures the quality of the annotations, and recall, which measures the quantity of annotations found.

B.2.1 Precision and Recall

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Sometimes, the language model (LLM) annotates divergently or makes syntax errors in our annotations, reducing its accuracy. Consequently, we take into account an annotation error rate (AER).

B.2.2 Kappa

$$\text{AER} = \frac{\text{True Positives} + \text{False Positives}}{\text{Total Annotations}}$$

To evaluate the effectiveness of the model, we use Fleiss' Kappa (k), a statistical measure of agreement among multiple annotators, including the LLM, which accounts for chance agreement. With P as the proportion of observed agreements and P_e as the proportion of agreements expected by chance :

$$\text{Fleiss' Kappa (k)} = \frac{P - P_e}{1 - P_e}$$

We categorize the values of Kappa into three levels, weighted with AER :

$$E_y = k \times (1 - \text{AER})$$

— Low : $k < 0.20$

— Moderate : $0.21 < k < 0.60$

— High : $k > 0.60$

B.3 Effect of Gender

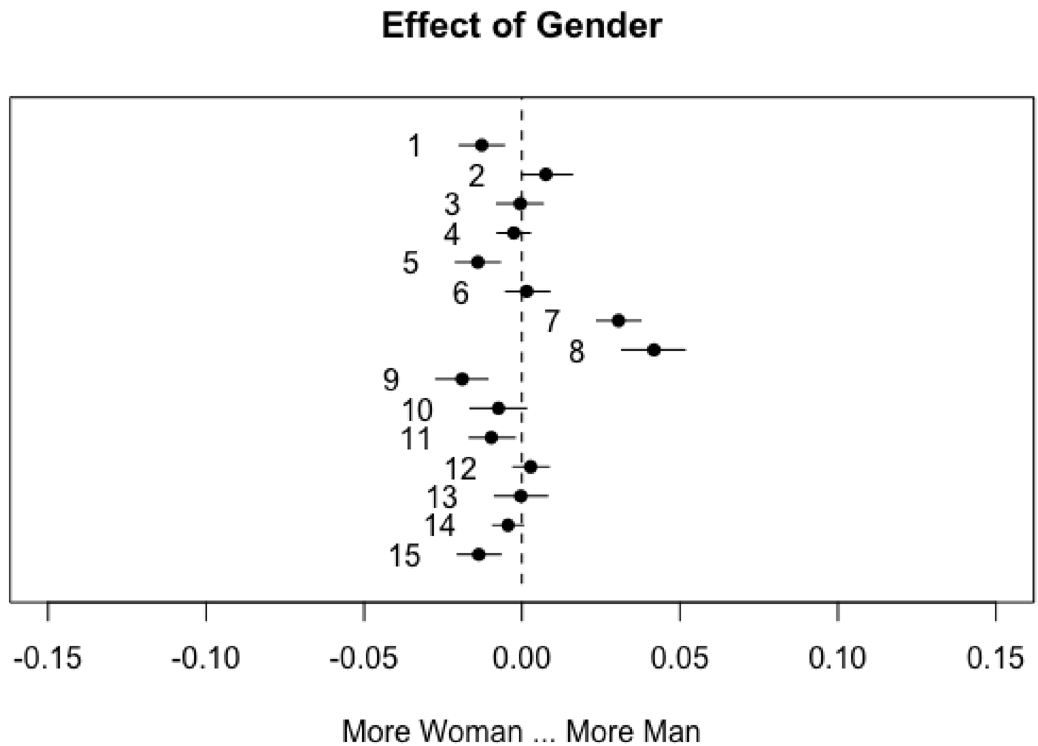

FIGURE 8 – Effect of Gender