The Life Cycle of Ideas*

Philippe Aghion Antonin Bergeaud Luc Paluskiewicz
Gaétan de Rassenfosse Raphaël Wargon

This version: July 13, 2026.
Latest version

Abstract

We contribute to the debate on the production of ideas by examining the scientific output of an entire field. We map the evolution of economic research by classifying 300,000 articles published in 1950–2020 using text analysis methods. We identify 90 distinct topics, each characterized by a coherent set of themes and vocabulary. We document three main empirical patterns. First, topics tend to show a rise-and-fall life cycle: their publication share increases during an expansion phase, peaks, and then gradually declines. Second, despite these topic-level cycles, the aggregate scientific activity grows over time, both at the global and topic levels. This growth does not come at the expense of quality. Third, this growth is accompanied by a constant "reset" in the topic composition of the discipline, with newer topics progressively displacing older ones. Altogether, our findings have implications for the debate about whether or not ideas are harder to find.

*We wish to express our sincere gratitude to Gustave Navez, Florian El Gourdou and Souleymane Faye for excellent research assistance. This work is supported by the European Research Council under Grant No. 786587, Proposal No. 101198463-BTI.
Addresses: Aghion: Collège de France, INSEAD, London School of Economics; Bergeaud: HEC Paris, Collège de France, CEP-LSE, CEPR; Paluskiewicz: Paris School of Economics, Collège de France; de Rassenfosse: École Polytechnique Fédérale de Lausanne; Wargon: Paris School of Economics, Collège de France. Email: luc.paluskiewicz@psemail.eu (corresponding author).

1 Introduction

In their 2020 book *Good Economics for Hard Times: Better Answers to Our Biggest Problems*, Abhijit Banerjee and Esther Duflo contrast Robert Gordon's pessimistic vision of innovation and long-term growth with Joel Mokyr's optimistic view. This discussion informs the debate on secular stagnation and the associated "ideas are harder to find" argument.

On the one hand, Gordon (2000, 2017) argues that recent technological revolutions (in particular the IT revolution) may be less transformative than earlier general-purpose technologies such as the steam engine, electricity, and the combustion engine. A complementary mechanism emphasizes that the marginal cost of pushing the frontier rises as knowledge accumulates. Jones (2009) formalizes the "burden of knowledge," while in the same vein Bloom et al. (2020) defend the idea of a structural decrease in the productivity of research: an increasing number of researchers are needed to achieve a given level of productivity growth or a given flow of innovation. The authors also look in detail at specific sectors, notably semiconductors, agriculture, and health, providing evidence that ideas, indeed, seem "harder to find." These forces feature prominently in modern R&D-based growth frameworks designed to eliminate scale effects and generate semi-endogenous growth (e.g., Jones 1995; Segerstrom 1998; Dinopoulos and Thompson 1998).

On the other hand, Mokyr (2014) stresses the continued arrival of major breakthroughs, and Mokyr (2018) argues that growth can be sustained by the interaction between scientific advances, recombination, and continued improvements along existing lines. More recently, Aghion et al. (2025) develop a model in which development opportunities are exhausted on each individual product line, but where the eventual exit of incumbents opens the door to new product lines, thereby "resetting the innovation clock." Their paper builds on Aghion and Howitt (1996) by distinguishing between research, which leads to the discovery of new product lines, and development, which consists of incremental improvements to existing product lines. Furthermore, recent micro-evidence suggests that flat aggregate growth alongside rising research effort need not reflect declining marginal research productivity: using U.S. Census microdata, Ando et al. (2026) estimate that the output elasticity (and gross marginal returns) of firm R&D in manufacturing have risen since the late 1970s, while faster obsolescence—linked to intensified technological rivalry—reduces the accumulation of effective knowledge and dampens aggregate productivity growth.[1]

We contribute to this debate by examining the scientific production of an entire field. Specifically, we

[1]This mechanism resonates with evidence of rising patent-output elasticities but weaker translation of patenting into firm growth (Fort et al., 2025), with findings of time-varying R&D depreciation (Li and Hall, 2020), with work documenting changes in spillovers and rivalry (Bloom et al., 2013; Lucking et al., 2019), and with firm-level evidence against simple fishing out at the knowledge-production stage (Klüppel and Knott, 2023).

study the rise and fall of topics in economic research, seeking to characterize the production of ideas. We develop a machine-learning method that clusters economics publications based on their content, and interpret clusters as meaningful "topics" in economics. Our topics behave in a similar fashion as the specific sectors in Bloom et al. (2020). The empirical analysis relies on the abstracts of over 300,000 journal articles published from 1950 to 2020. Our focus on economic papers helps us validate the results using our expert knowledge, in particular to assess the meaningfulness of the topics generated by our algorithm. Specifically, we implement a machine-learning algorithm that performs unsupervised, text-based cluster identification. Our method fixes the number of topics throughout the study period, though topics can be more or less active over time. Each paper, in turn, relates to a mixture of topics and the relationship between a paper and the various topics is based on the embeddings of words appearing in the paper's abstract, following Dieng et al. (2020). We let the number of topics emerge endogenously based on algorithmic performance metrics, namely out-of-sample performance and analytical coherence. This process leads us to consider 90 topics, each corresponding to a specific object or method of study. Having allocated all the papers in our dataset to these 90 topics, we proceed to studying the evolution of publications and citations within each topic over the full period.

We report three main findings. First, topic importance tends to follow a bell-shaped curve. We proxy a topic's importance using the share of articles that relate to it among the total number of new articles in a given year. We also find that for a topic that follows this typical trajectory, it first catches up to the yearly average production by pre-existing topics as it gains importance, then falls below the average progression of the entire economic field. This pattern echoes Gordon's "fruit tree parable"—research within a topic first harvests the low-hanging fruit, after which remaining ideas become progressively harder to find. Second, the overall production of economic publications and citations displays sustained growth throughout the period. That is, the aggregate production of economic publications exhibits no diminishing returns. Third, we find a perpetual "reset" in the production of economic ideas, with the constant emergence of new topic combinations. This pattern, in turn, is consistent with the absence of a decline in the aggregate growth of the number of publications, aligning with Mokyr's more optimistic view.[2]

The remainder of the paper is organized as follows. Section 2 presents the data and outlines our method. Section 2.3 discusses the results of the clustering algorithm. Section 3 presents our main findings. Section 4 concludes. The data acquisition and analysis pipeline is available in open access to encourage

[2]Our topic life-cycles also connect naturally to the growing "science of science" literature documenting slowing disruptiveness and increased canonization as fields expand (e.g., Wuchty et al. 2007; Uzzi et al. 2013; Wu et al. 2019; Chu and Evans 2021; Park et al. 2023; Foster et al. 2015).

reproducibility.[3]

2 Data and Method

2.1 Data Acquisition

We collect data on publications and their abstracts from OpenAlex, a bibliometric database launched in 2021 that uses open-access data from Microsoft Academic Graph (MAG) (Priem et al., 2022). The OpenAlex database contains about 250 million academic works and 90 million researcher profiles. We focus on the period between 1950 and 2020 during which OpenAlex is exhaustive.[4] In addition to being fully open access, OpenAlex is also the database that covers the largest number of journals and articles, offering another advantage over alternatives such as the Web of Science (WOS) and Scopus (Visser et al., 2021; Alperin et al., 2024; Culbert et al., 2025). The fraction of papers with available abstracts is slightly lower in OpenAlex than in WOS, reaching 87 percent and 92 percent, respectively. However, this figure is acceptable as omitted abstracts are unlikely to bias our analysis. Moreover, all researchers who have a profile in WOS and Scopus also have a profile in OpenAlex (Zhao and Chen, 2025).

OpenAlex automatically classifies publications into scientific fields (and topics within fields) using a structural topic modeling algorithm, which itself relies on the Scopus ASJC classification.[5] This process does not incorporate expert knowledge and spans the entire database, resulting in articles in management and finance often being categorized as economics articles. Conversely, economics articles may be classified in other fields. For instance, mechanism design is considered as a subfield within the broader field of decision science.[6] Because our analysis critically relies on properly scoping economic research articles and identifying topics within the broader field of economics, we will perform our own identification of economics papers and deploy a dedicated classification algorithm to identify topics.[7]

Identifying economic papers

Delineating the scope of papers that count as "economics" is a surprisingly challenging task. One could consider only papers published in the Top 5 journals or the Top 20—assuming we agree on such a

[3] The code will be available upon publication of the article.

[4] MAG was discontinued in 2021, so that the number of available sources for OpenAlex drops after that point (Alperin et al., 2024).

[5] See the official OpenAlex topics documentation and the relevant GitHub repository.

[6] See Section E.

[7] The well-known JEL codes are not well-suited to our analysis. They are assigned at the time of publications and may, therefore, miss nascent topics. Furthermore, they are assigned inconsistently and potentially subject to strategic manipulation.

list—but this selection would lead to a very restricted number of papers, potentially missing important patterns (*e.g.*, the fact that new fields may emerge or decline in less prestigious journals). This approach would also require the list of top journals to be stable throughout the decades. An alternative approach is to start with the list of all journals in which well-recognized economists, such as Nobel Prize winners, have published. However, this list would include journals that are not core to the economics discipline, such as John Nash's 1951 *Annals of Mathematics* paper, Daniel Kahneman's 1974 *Science* paper, or Elinor Ostrom's 2007 *PNAS* paper. While these selected articles may be relevant to the field of economics, the vast majority of articles in these journals would not be. One could also consider all 4000+ journals listed in RePEc or the 700+ journals in EconLit, but such a list would include journals such as *Conflict Management and Peace Science* or *Foresight* which, we would argue, are not core to economics.

We build our list of journals from curated lists. Specifically, we start from the list established in 2020 by the *Centre National de la Recherche Scientifique* (CNRS) — the French equivalent of the U.S. National Science Foundation. Within that list, we focus on those journals that are categorized as general or specialized economics. We exclude journals categorized in "health economics" due to the large overlap with public health journals. This selection leads us to consider 359 journals. To that list, we add the top 300 journals in the 2025 SciMago list of best ranked journals in Economics and Finance.[8] Among those, we only keep economics journals, discarding journals in finance. This source leads us to add 54 journals to the list. We then restrict the list to the subset of journals that are available in OpenAlex (7 journals not retained). Finally, we remove journals that are not in English (50 journals removed). We obtain a final list of 356 journals,[9] representing 554,224 publications in OpenAlex.

Data preparation and cleaning

We retain publications that have an abstract and a valid year of publication, leading us to consider 457,660 papers published between 1950 and 2020.[10]

Each publication in our resulting database is then associated with a unique identifier and a raw text, corresponding to the paper's abstract. We then perform additional cleaning to improve the performance of our clustering algorithm. Specifically, (1) we filter out papers for which the abstract is not meaningful (96,417 abstracts removed);[11] (2) we remove elements from the text that are likely to bias our analysis:

[8] The CNRS list and the SciMago list are both available online.

[9] Some journals have several identifiers in OpenAlex.

[10] While all the selected journals published the abstracts, it is occasionally missing for some publications. Furthermore, we exclude publications that have an abstract of less than 100 characters.

[11] For example, when the text is clearly a webscraped text indicating how to navigate the website instead of an abstract.

plain-text citations in the abstract, information about the journal of publication or JEL codes, which are sometimes appended to the abstract; (3) we perform part-of-speech tagging in order to remove stop words, keeping only adjectives, adverbs, nouns, and verbs. We also remove verbs and adverbs that lie in the top 0.1% of frequency distribution, meaning that they are present in a large number of abstracts, independently from the topics the publication is associated with; (4) following Fan et al. (2017), we remove words from a hand-curated set of "corpus-specific stopwords" (such as "paper," ...), and country-related terms (e.g., "United States," "French," "American," etc.) to only keep meaningful keywords; (5) we stem the words in each abstract to facilitate grouping related terms. Stemming reduces a word to its root form (e.g., "development" and "developing" are both mapped to "develop").

The left panel of Figure E.3 presents the numbers of publications and author profiles in our sample from 1950 to 2020. Both series are increasing at a high rate over that period: economics papers have grown at an annualized rate of 3.2 percent since 1950 whereas the yearly number of publishing authors grew by 4.7 percent over the same period. The increase in the number of articles over time is partly driven by improvements in data availability; however, it also reflects a genuine increase in the overall scientific production of economics papers. That the number of authors rises much faster than the number of publications in our sample may seem to point to decreasing returns to research in economics, echoing the viewpoint of Jones (1995) and Gordon (2017). On the other hand, there may be other explanations for this specific trend, for example a widening of the set of authors publishing in the journals sampled. If these journals open up to a broader set of authors over time but these authors are also publishing in other journals, then we should expect the in-sample number of authors to rise faster than the number of papers. The right-hand panel of Figure E.3 confirms this intuition. It shows that the average number of publications of the authors who publish in our curated list of journals rises over time when considering all their articles recorded in OpenAlex (i.e., also those that we did not classify as economics).[12,13] Yet, the production of papers published in our core list of journals remained stable, as the bottom line of the right panel indicates. Put differently, researchers in our sample are increasingly publishing in a more diverse set of journals—a mechanical implication of capacity constraints once we fix the number of journals.

2.2 Allocating Papers to Topics

Having assembled a large corpus of abstracts spanning the economics discipline, the next step involves grouping papers into interpretable topics (or "clusters"). We do so by estimating an Embedded Topic

[12]We weigh each publication in this graph by the number of coauthors to account for the increasing numbers of co-authors on economics papers over time.

[13]We further discuss the gap between the sample and the rest of the authors' publication in Section E.

Model (Dieng et al., 2020) on a 20% training subset of the data. We then select the model that best fits the data by looping over candidate specifications (including the number of topics and other tuning parameters) and comparing their performance using a set of quantitative metrics. Once the preferred specification is identified, we apply it to the full corpus to obtain a final many-to-many allocation of papers to topics. This section provides an intuitive overview of the procedure; Appendix F presents the technical details. We first explain how topics are defined and how papers are assigned to topics for a fixed, exogenous number of topics. We then describe how we choose the number of topics endogenously.

Method to identify topics

We model each paper as a mixture of topics, such that each paper loads on multiple topics with different weights. For example, an abstract may combine a "growth" component with a "dynamic modeling" component, reflecting the presence of vocabulary associated with each theme. The key identifying variation comes from patterns of word co-occurrence across abstracts, i.e. words that repeatedly appear together tend to define the same topic. In a growth-related topic, for instance, terms such as growth, productivity, and develop will tend to co-occur, whereas terms unrelated to that theme will not.

A practical complication is that semantically close terms need not share the same surface form (e.g., female vs. women), and treating them as unrelated can fragment otherwise coherent topics. To address this point, we represent text using embeddings that capture semantic proximity, drawing on the SPECTER model (Cohan et al., 2020). Intuitively, abstracts (and the terms they contain) that are used in similar contexts are mapped to nearby points in the embedding space. In contrast to a pure bag-of-words representation that only counts token occurrences, an embedding-based representation allows the model to exploit semantic similarity when forming and interpreting topics.

Next, fix a number of topics K. For each paper j, the model estimates a vector of topic shares $\theta_j = (\theta_{j1}, \ldots, \theta_{jK})$, where $\theta_{jk} \geq 0$ and $\sum_{k=1}^{K} \theta_{jk} = 1$. Each topic k is characterized by a topic–word distribution β_k, with entries $\beta_{kw} = P(w \mid k)$ indicating how strongly word w is associated with topic k. A paper with a high weight θ_{jk} is, therefore, more likely to contain words with high $P(w \mid k)$. Equivalently, the model represents the distribution of words in paper j as a mixture,

$$P(w \mid j) = \sum_{k=1}^{K} \theta_{jk} \, P(w \mid k). \tag{1}$$

For example, if paper j loads heavily on an "auction theory" topic, the word *auction* will have a high probability under that topic—though it may still appear with lower probability under other topics. The Embedded Topic Model combines information from word frequencies with embedding-based

representations that capture semantic similarity when estimating these distributions (Dieng et al., 2020).

Given initial values for $\{\theta_j\}_j$ and $\{P(w \mid k)\}_k$, the algorithm evaluates how well the model explains the observed abstracts using a likelihood-based objective. It then iteratively updates the topic proportions across papers and the word distributions within topics to improve this fit, converging to the set of topics and paper-specific topic shares that best explain the data, conditional on K.

Method to identify the optimal number of topics

As an illustration, in the upper panel of Figure 1, we show different grouping possibilities for five papers with a fixed number of groups, $K = 2$. Let us imagine that the 2-dimensional distance between two points represents their semantic difference—this is what we do when using embeddings. If there are more important words found in A, B, C, and D than in A, B, and E or C, D, and E, then the configuration in the right panel will be the one that maximizes the likelihood (in this example, it is the configuration which minimizes the distance between papers inside of a topic).

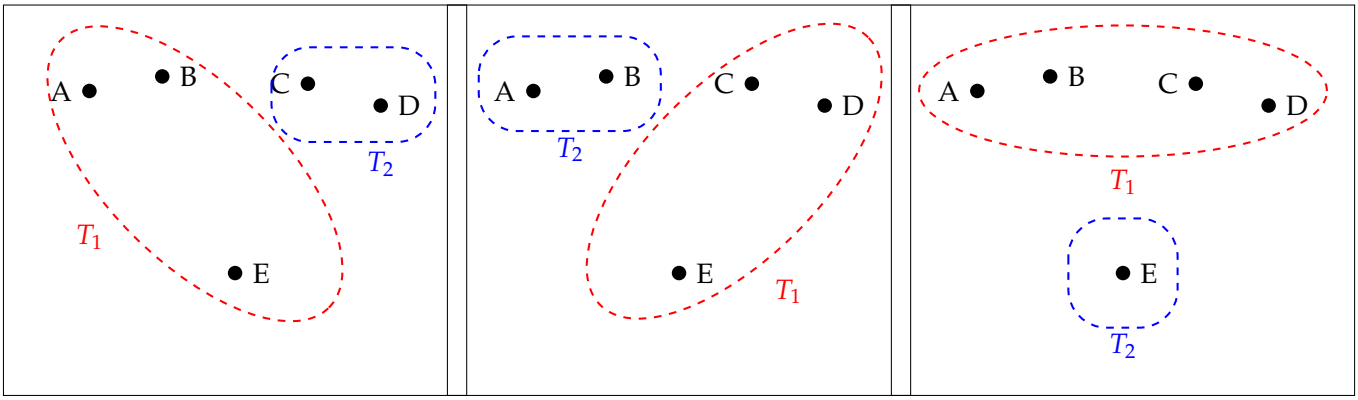
Next, we let the number of topics K vary and we assume a fixed cost per topic in order to compute the optimal number of topics K^*. We provide an illustration of this process in the lower panel of Figure 1. In this example, we have a trade-off between minimizing the distance between articles within a topic and the total fixed-cost. If the cost per topic is sufficiently high, we will prefer to choose $K = 3$, i.e., the middle graph.[14]

Next, we allow for new papers to come in. We want to know which topic structure will be most robust to introducing new papers. We expect those structures that pass this robustness test to display a small number of topics with a high number of closely related papers within topics. Such robustness, the inverse of which we will refer to as perplexity, may stop decreasing after the optimal number of topics is reached. With two topics, the distance between new papers and the center of the closest topic will be high. It diminishes when using $K = 3$ but stops decreasing at $K = 4$.

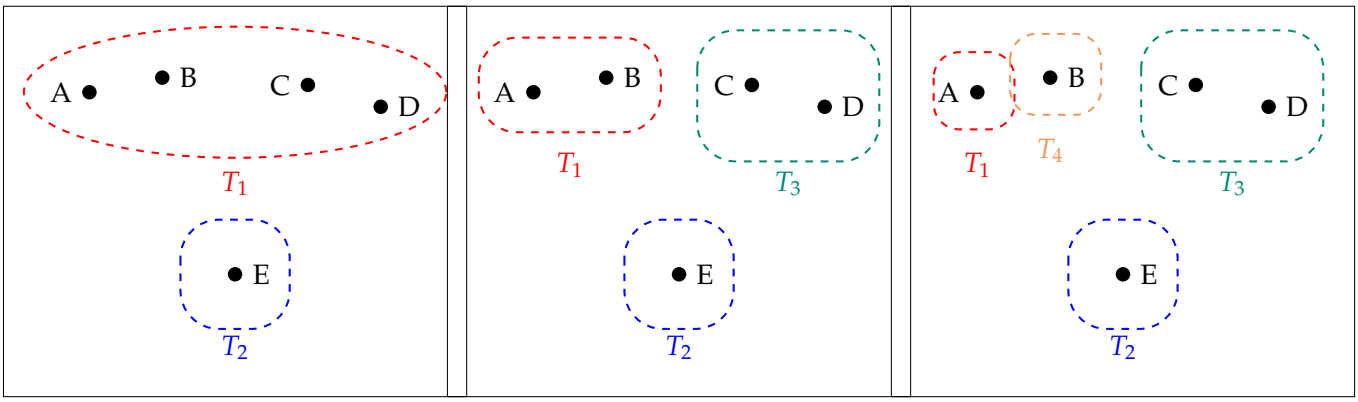
More generally, we use a metric based on the notion of "perplexity" to evaluate the performance of each estimated model. Perplexity measures the extent to which a model is "surprised" by new data it has not seen before. It is an inverse measure of the likelihood performance of the model. A limitation of the topic selection method based only on perplexity to determine the optimal number of topics is that the topics generated by this method may be hard to interpret. For that reason, we also use an alternative "coherence" metric based on semantic distance, i.e. on the semantic similarity between the top words

[14] In practice, in our approach, the cost represents the increase in perplexity generated by the model when adding another topic, in other words the decrease in the likelihood of the model fitting the data when increasing the number of clusters.

Figure 1. Topic Allocation Strategies



(a) Three different valid allocations for $K = 2$



(b) Optimal allocations for $K = 2$, $K = 3$, and $K = 4$

Note: Each sub-figure displays five papers (A, B, C, D, E) represented as points in a semantic space. Coloured dashed regions delimit papers assigned to the same topic T_k. **(a)** for a fixed $K = 2$, multiple valid allocations exist — the grouping is not unique, and different partitions can yield very different levels of semantic coherence. **(b)** as K increases from 2 to 4, each paper is assigned to an increasingly fine-grained topic, improving within-topic homogeneity at the cost of a larger and potentially less interpretable model.

within a topic, as measured by the co-occurrence of these words in the data compared to how high up they rank within the topic. We provide more detailed explanations about the corresponding algorithm and validation metrics in Appendix F.

Our final algorithm uses a three-dimensional optimization procedure that jointly optimizes coherence and perplexity over a grid defined by three parameters: the minimum document frequency and the maximum document frequency of the words included in the model, and the number of topics. More specifically, we compute the values for coherence and perplexity for each of the points in the grid. We then select the model that produces a maximum and/or a kink in the variation of the corresponding metrics outcomes. We check the robustness of this approach by applying it to simulated data. The results

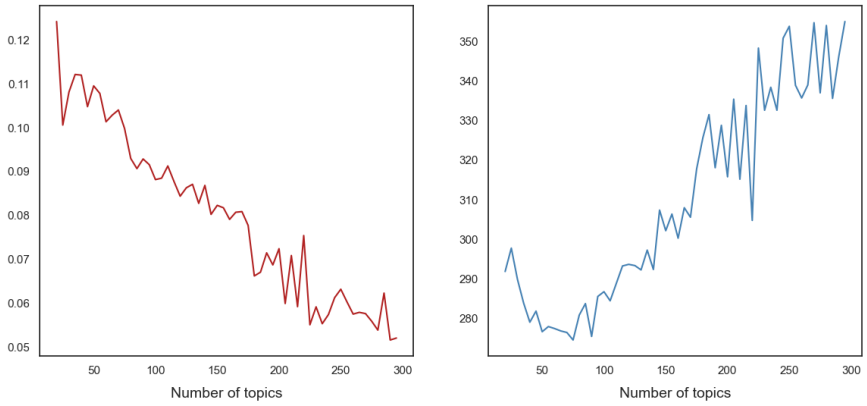
of this exercise are presented in subsection G.2.

2.3 Implementation of the Algorithm

The optimization procedure leads us to retain a minimum document frequency of 1% and a maximum document frequency of 50% (see Section G for details). In other words, we keep words that occur at least in 1% and at most in 50% of the abstracts. Given these two parameters, we then vary the number of topics and focus our attention on the evolution of the coherence and perplexity metrics.

Figure 2 presents the evolution of both metrics. On the left panel, coherence sharply deteriorates between 70 and 100 topics, while perplexity reveals a kink in the same region on the right panel. We set the number of topics to 90, but note that our analysis is robust to choosing 70 topics.[15]

Figure 2. Coherence and Perplexity, by Number of Clusters



Notes: The graph above presents topic coherence (left) and perplexity (right) as a function of number of topics chosen for the algorithm. The metrics are obtained from the same sample of 20% of the dataset for Embedded Topic Models varying only in the number of topics with minimum document frequency equal to 1% and maximum document frequency equal to 50%. The train/test split is 80/20 and is the same for all models. The model is estimated for a range of 30 to 300 topics, in steps of 10.

As explained in the previous section, each paper is allocated to a mixture of topics. For instance, our model allocates "On the mechanics of economic development" (Lucas, 1988) to 37% of topic #25, labeled "Economic Growth", and 3% of topic #46, "Dynamic Stochastic Models".[16] We can use these data to build a weighted sum of papers per topic. That is, if we sum the number of papers belonging to topic #25, this paper will count for 0.37 within that topic. In this example, we see that the Lucas paper mostly relates to one topic, with all other values being in the low single digits. A value of 3% is not negligible in this

[15]Results are available upon request from the authors.
[16]The index number assigned to each topic is arbitrarily chosen by the ETM algorithm and does not reflect importance.

context: under a uniform distribution across topics (which a random allocation would produce), each topic would have a weight of $1/90 \approx 1.1\%$. Thus, 3% is more than twice the uniform random expectation. Note also that the model assigns a positive weight to each of the 90 topics, with weights arbitrarily close to 0 for less-related topics. Some of the variation generated by this method is, therefore, simply noise.

In order to filter out this noise, we truncate the distribution to the n^{th} topic with the highest value and then reweigh the distribution based on these shares alone. In our example, limiting at $n = 2$, we would have topic #25 weighting $37/(37+3) \approx 92\%$ and topic #46 weighting about 8%. Throughout the paper, we set $n = 10$, unless otherwise noted. Finally, we also define the *importance of topic k in year t* as the share of papers related to k in t, which we compute as the ratio between the unweighted number of papers that include topic k in its first n^{th} topics and the total number of publications in t.

3 Stylized Facts on the Lifecycle of Ideas in Economics

Having allocated papers to topics, we now seek to characterize the lifecyle of topics. We present three main empirical findings: *(i)* individual topics tend to follow a bell-shaped life cycle; *(ii)* there is overall a global growth in papers; and *(iii)* new topics consistently emerge, displacing older ones.

3.1 The Topic Life Cycle: Evidence of a Rise-and-Fall Pattern

The measure of importance described above allows us to derive the trajectory of each topic in the economic literature. We use these trajectories to test Gordon's argument that once a subject is discovered, new ideas related to this subject become scarcer. For each topic, we fit a Locally Estimated Scatterplot Smoothing (LOESS) curve on the trajectory of importance, estimated here as the number of publications in which this topic is among the 10 most important.[17] Our full results are shown in Figure A.2. A large majority of the trajectories (62 topics) can be classified as bell-shaped and single-peak; the shape that would be expected if ideas indeed become harder to find for each new Kuhnian discovery.[18] Among the remaining topics, five appear to be continuously decreasing and a further eight continuously growing—possibly representing bell curves that have already peaked and not yet peaked, respectively. The remaining topics that appear to regain importance during the period after declining, have a pattern that is not consistent with the hypothesis that ideas become systematically harder to find in each topic.

[17] This process is detailed in Section C.1.

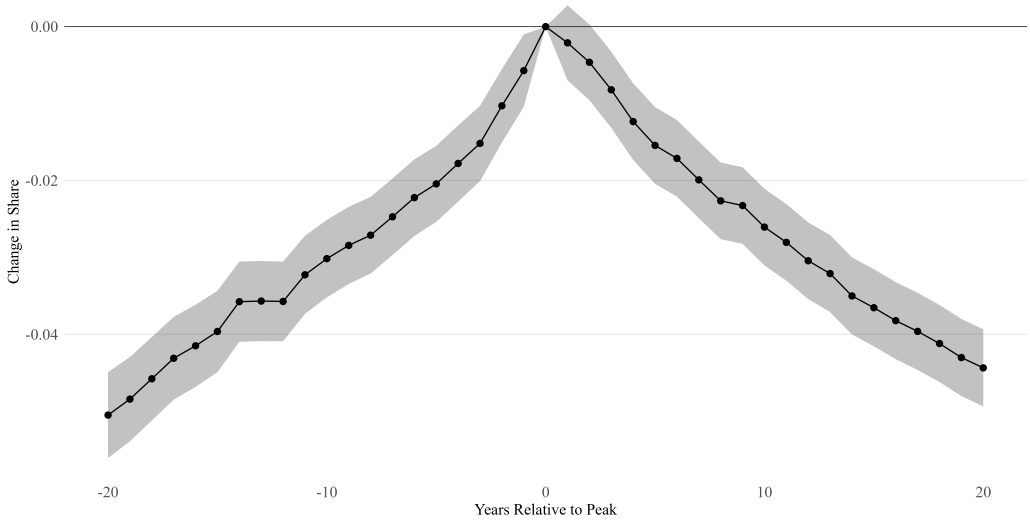
[18] The philosopher and historian of science Thomas Kuhn introduced the idea that scientific progress occurs through paradigm shifts—major breakthroughs that open new scientific paradigms or research trajectories.

For the subset of topics that we identify as bell-shaped, we seek to characterize the "typical" evolution of topic importance around the peak. For each topic k, let τ_k^* denote the year in which its publication share is maximal, and let $y_{k,\text{peak}}$ be that maximal share. We define event time $t = \tau - \tau_k^*$ and express topic importance as the ratio of the topic's share in year τ to its peak value $y_{k,\text{peak}}$, so that $y_{k0} = 1$. We then estimate the average event-time profile by running the following event-study regression:

$$y_{kt} = \sum_{s=-20}^{20} \beta_s \mathbb{1}\{t = s\} + \alpha_k + \varepsilon_{kt}, \tag{2}$$

where y_{kt} is the (smoothed) normalized publication share of topic k at event time t. The publication share is smoothed using a three-year trailing average (year τ and the two preceding years).

Figure 3. Average Topic Popularity Trend



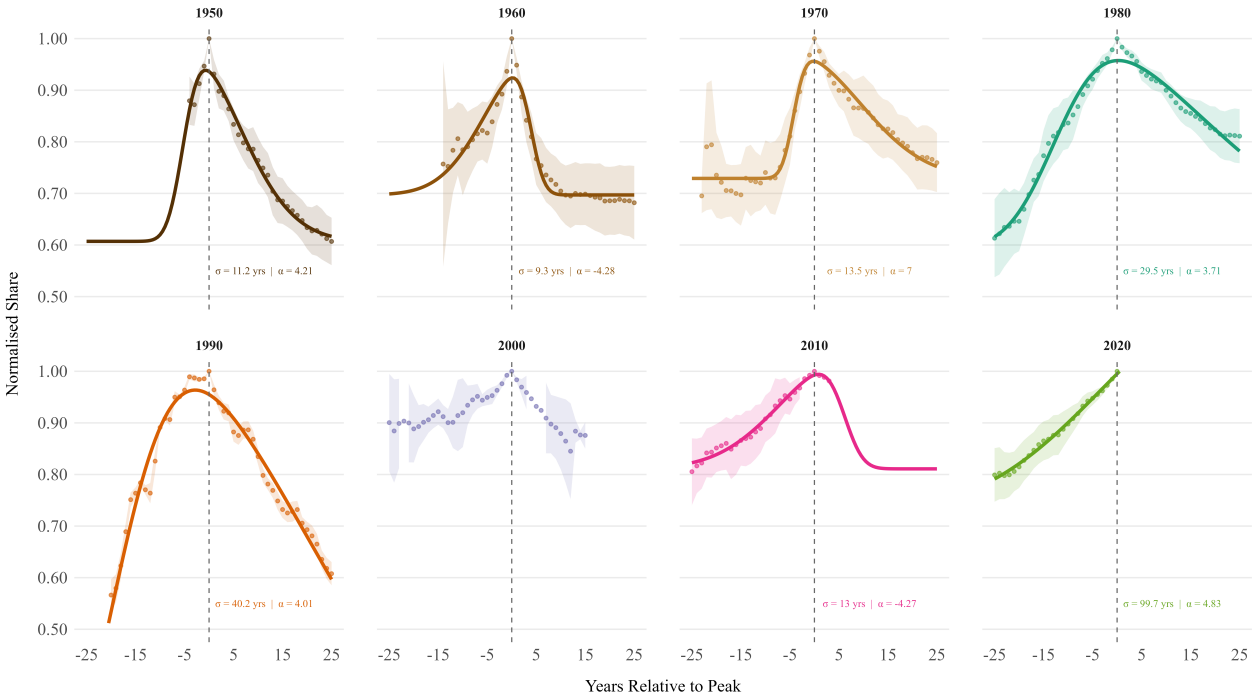
Notes: The graph above presents the average share of publications per topic relative to the year in which the maximum is obtained, normalized by the topic-specific average. The share of publications is computed as the share of publications which have this topic in the 10 highest values of their distribution per year.

Figure 3 plots the coefficients β_s and the 95-percent confidence intervals. The variation in the share of publications belonging to a given topic with a bell-shaped trajectory relative to its peak is quite substantial. We find that on average, 20 years prior to its peak, the share of papers allocated to a given topic is 4 percentage points (pp). below the peak, a little less than four times the aforementioned uniform random share of $1/90 = 1.1$pp.

Figure 4 further suggests that the bell-shaped pattern looks similar across time, independent of the decade in which a topic reaches its peak. The figure shows the evolution of the share of papers belonging to a given topic relative to its peak, distinguishing by the decade in which the peak is reached. We see no

clear difference in the shape of the bell curve between topics peaking in recent years and those peaking before the 1980s. This illustration further shows that, today, topics are not at the same stage of their life cycle; topics that reached their peak in 2020 seem to be in the first phase of their cycles, while topics that reached their peak previously are monotonically decreasing.

Figure 4. Share of Publications per Topic, by Decade of Peak



Notes: The graph above presents the average share of publications per topic relative to the year in which the maximum is obtained, normalized by the topic-specific average, split by decade of peak. The share of publications is computed as the share of publications which have this topic in the 5 highest values of their distribution per year.

3.2 Global Growth in Economic Ideas

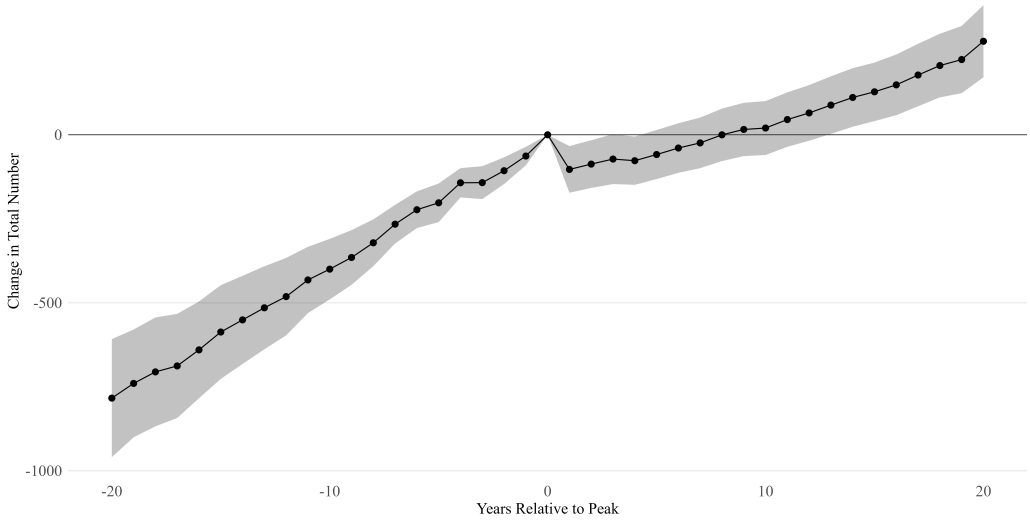
The descriptive statistics in Figure E.3 depict a steady growth of publications over time. We now seek to shed light on the sources of this growth.

Topics only decline in relative importance, not in absolute terms

This time, we no longer use the relative measure of topic importance; instead, we rely on the (weighted) number of publications in a given topic. Figure 5 shows the β_s coefficients associated with a model where the dependent variable is the average number of publications per topic relative to the year the maximum is reached, controlling for the topic fixed-effect. Unlike the previous illustrations, which showed a clear monotonic decline in topic (relative) importance once the peak is reached, the total number of publications

in a topic continues to increase even *after* the importance peak is reached.[19] Put differently, while topics can eventually become less prominent in relative terms, the overall volume of publications per topic continues to increase.

Figure 5. Average Topic Trend



Notes: The graph above presents the average number of publications per topic relative to the year in which the maximum is obtained, normalized by the topic-specific average. The share of publications is computed as the share of publications which have this topic in the 10 highest values of their distribution per year.

No evidence of quality decrease across a topic's lifecycle

Yet, focusing on publication quantity alone provides only a partial view of output. Not all publications are equal in impact and novelty. We now turn our attention to measures of publication quality. Did the increase in publications come at the expense of quality, with more marginal contributions over time? Three pieces of evidence suggest that there is no substantial change in publication quality as topics mature, and eventually fade, in relative importance.

First, we measure quality using the number of citations a paper receives up to 3 and 5 years after publication, to avoid potential truncation bias.[20] Figure D.1 shows the average total citations received by publications in a topic in the three and five years following their publication, by year, normalized by the topic-specific average. This average number of citations received by topics increases each year. Growth

[19] As explained in subsection 2.1, we only capture a subset of the publications of these authors in the data – the apparent diminishing returns observed in sample are just a byproduct of our selection process.

[20] We obtain similar results without truncation and truncated at the 1-year level.

in "quality" is, however, different from growth in quantity in that it does not follow the same bell-shaped pattern relative to the peak of importance of a topic. However, this growth in the average number of citations received could be an artifact of the increase in the pool of citing papers over time.

Second, we use the originality score by Hall et al. (2001), which is a measure of the dispersion of topics on which a paper builds (*i.e.*, that a paper cites).[21] We also use the generality score, which computes the dispersion of topics building on this paper.[22] Figure 6a depicts the global evolution of both scores over time. The point estimates represent the coefficients β_s in a model where the dependent variable is the average of papers' originality/generality scores in a given topic-year. Originality increases monotonically, which implies that, across topics, papers draw inspiration from a more diverse set of fields over time. Likewise, the generality score does not decrease even in the long term after the peak of the topic is reached, meaning that papers keep being used in a variety of other topics.

Finally, we compute a measure based on sentence embeddings from Cohan et al. (2020) of similarity of a paper compared to its references and to the papers it references. We assume that the further away the paper is located from its cited papers, the higher its novelty, in line with Arts et al. (2025). Similarly, the closer the paper is located from the papers that cite it, the more influential it is. Figure 6b shows no decrease in either measure, further supporting the view that the increase in scientific output did not come at the expense of quality.[23]

3.3 Reset of the Innovation Clock

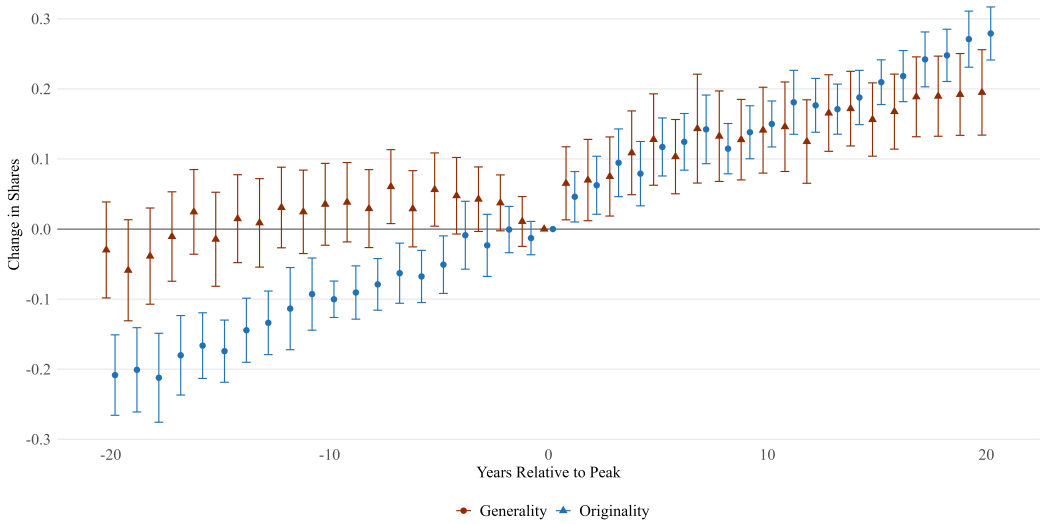
That topics only decline in relative importance but not in absolute importance is sufficient to explain the global growth in ideas. In this section, we further characterize aggregate growth, providing evidence of a constant reset of the "innovation clock." Figure 7 depicts the histogram of the number of new long-lasting combinations of topics by year, illustrating regular emergence of new combinations. A "new long-lasting combination" arises when two topics occur as first topic on publications for a period of at least ten years. We see no decline in the flow of new topic combinations each year, which in turn speaks to sustained innovation reset. Furthermore, the average number of papers for each new combination when it arrives appears stable at a rate of around 30 publications per year.

[21] This measure was first developed in the context of patented inventions, with more original inventions drawing on a more diverse set of technology classes. While it is affected by growth in the number of papers – a paper published in a more recent year has a wider set of papers to choose from when deciding on citations – it is still considered informative about the novelty of a paper.

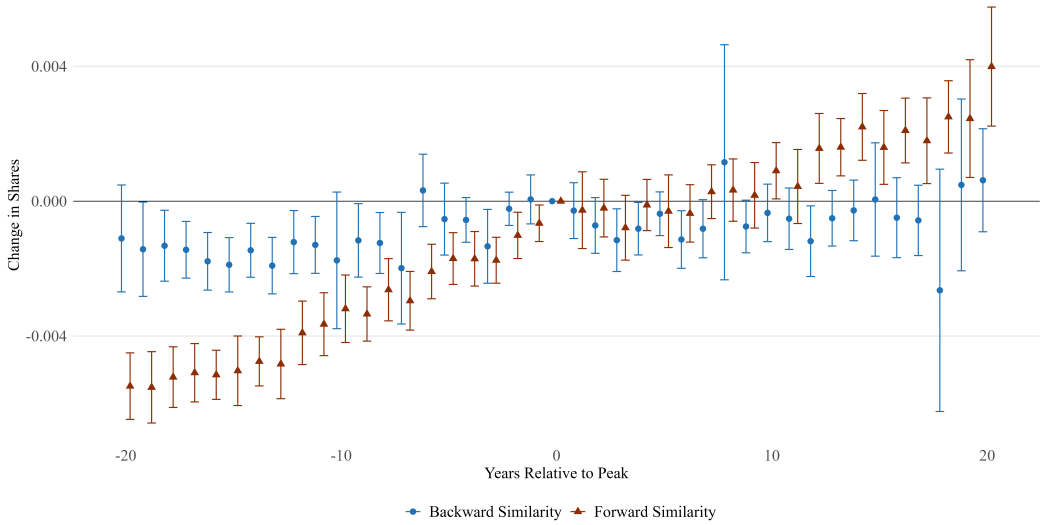
[22] Likewise, there is a larger set of papers that can cite a paper published in recent years due to growth in the total number of publications. We truncate the citations to build this score on the 5 years after publication of the focal paper in order to alleviate the bias, similarly to citations.

[23] We further discuss measures of novelty in Section H.

Figure 6. Average Topic Trend



(a) Average Topic Trend in Originality and Generality

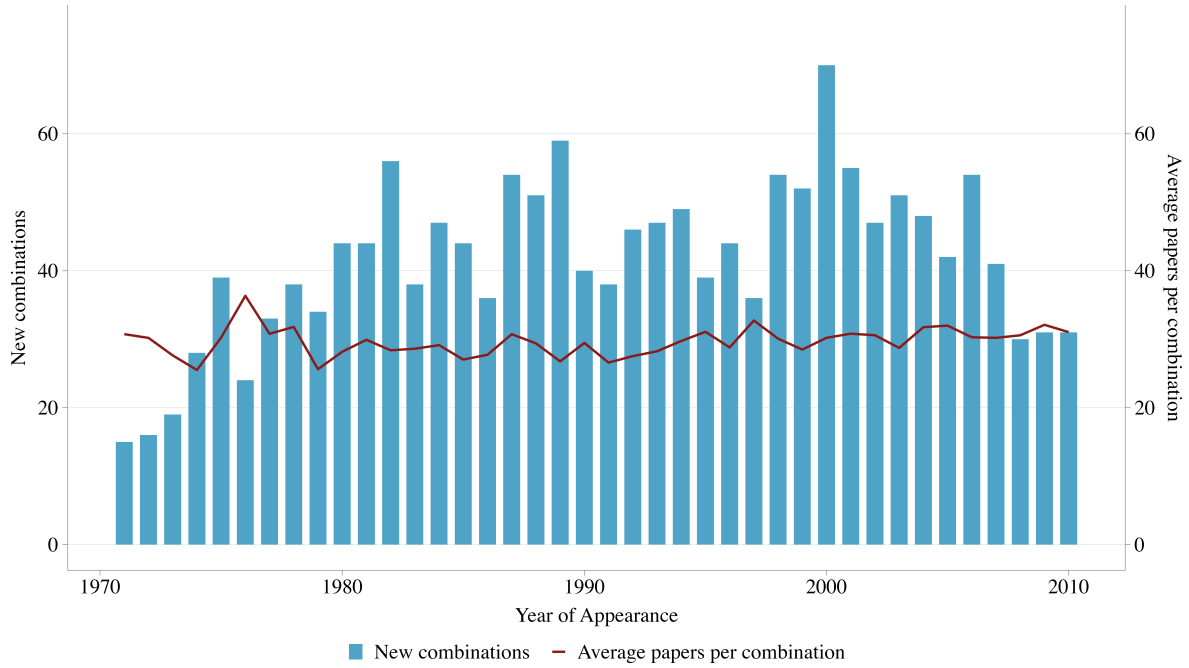


(b) Average Topic Trend in Similarity to Backward and Forward references

Notes: The graph above presents the average similarity to references and citing papers per topic, relative to the year in which the maximum is obtained, normalized by the topic-specific average. The graph above presents the average originality and generality (among papers citing in the next 5 years) per topic, relative to the year in which the maximum is obtained, normalized by the topic-specific average.

We also find that newer topics are more likely to generate new durable combinations. Figure 8 represents the average difference in birth years between two topics in a given combination: in red, the observed gap, in grey, when we randomly draw topics to generate simulated combinations based on

Figure 7. Number of New Combinations Emerging



Notes: The bar chart above presents the number of new combinations of topics per year of first appearance. We only take into account in this case the 2 first topics associated with one publication, independently of the order. In order to appear in this graph, combinations of topics need to have at least 10 years of non-null number of papers with the same combination. The line represents the average number of publications during that period for each new such combination.

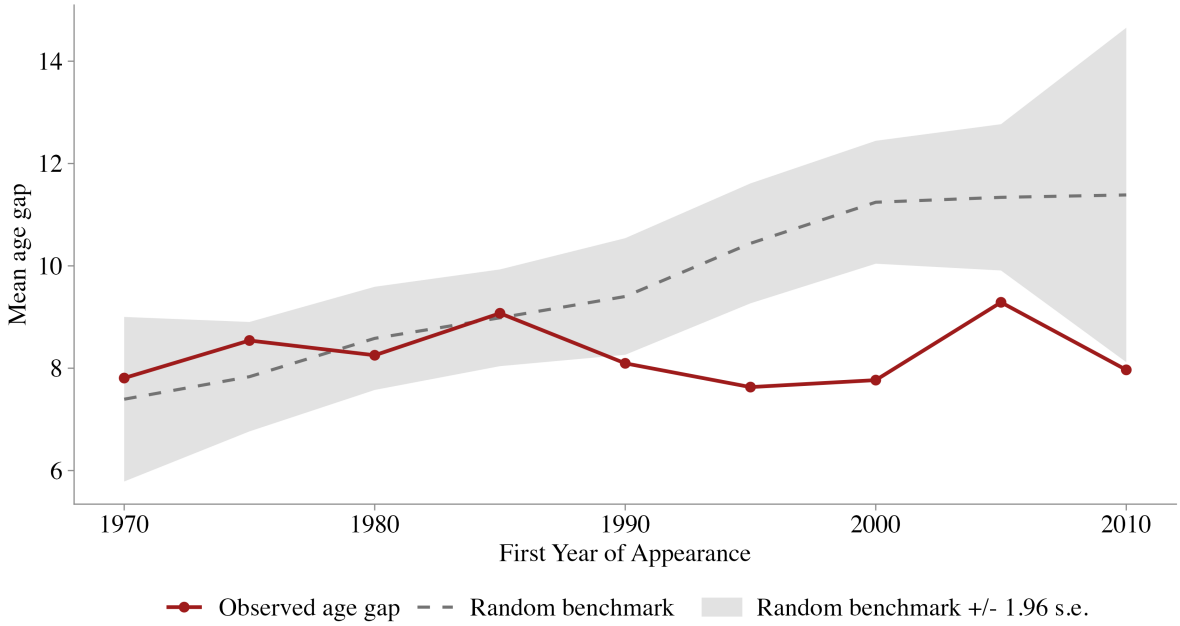
available topics each year.[24] We find that compared to a purely random process, the actual recombinations tend to be among younger topics. These recombinations supposedly form the seeds of new topics on which the economics discipline will continue growing.

New topics and quality

We now assess whether the resetting of ideas plays a role in the qualitative evolution of the discipline. Results in Section 3.2 suggest that quality does not follow the same bell-shaped curve as publication quantity. However, that does not mean that there are no differences in levels of quality when topics emerge. Figure D.2 shows that the topic-specific averages in our three measures of quality as computed for the aforementioned figures do not decrease on average over time of peak of the topic. This suggests that not only do newer topics keep appearing, but also that they have a higher quality on average compared to other topics. There are, therefore, two elements leading to quality growth. The first, which we have seen above, is that quality keeps increasing even in topics that have passed their prime. But

[24]Topics are drawn with replacement but not combinations. The trajectories are drawn 300 times in order to generate the standard errors.

Figure 8. Difference in birth years between two combined topics, observed and simulated



Notes: The red line represents the average difference in birth years between two topics in durable combinations appearing each five-year period in our data. The dotted grey line represents the same average value for randomly drawn combinations from average topics. Standard errors are obtained by drawing the combinations 300 times.

this could be due to the second element: new, higher quality topics take more space as the older ones experience a relative decline, leading to higher-quality papers in economics, and possibly pushing up the standards for papers remaining in topics that have passed their peaks.

4 Conclusion

In this paper, we contribute to one of the main research questions in innovation economics literature: *are ideas harder to find*? We implement a machine-learning method that generates clusters of publications that belong to the same topics and allocates all publications across these "topics." Our algorithm assesses the existence of 90 topics in economics between 1950 and 2020. Three main stylized facts emerge from our analysis. First, we show that the majority of topics follow a bell-shaped trend when looking at the share of publications within a specific topic, suggesting that topics—or research lines—seem to first rise before declining in relative terms. Second, we see a global growth in the aggregate number of publications, even for topics that have lost relative influence. Third, we find evidence for the emergence of new topic combinations that seem to occur at a constant rate. Overall, we show that ideas become eventually harder to find within each individual topic in economics, but not at the aggregate level.

Beyond contributing to the literature on research productivity, our paper also relates to a recent and influential literature on the evolution of economic publications. Krauss (2024) documents growth in the overall average number of co-authors per publication, and an increasing share of older authors. Angrist et al. (2020) use JEL codes to partition the aggregate production of economic papers across fields. They show evidence of an increasing share of applied microeconomics since the 1980s. Closer to our analysis and focus, Card and DellaVigna (2013) identify nine facts about the evolution of economic science over time. Using articles published in the Top-5 journals from 1970 to 2012, the authors classify papers into JEL-based fields. They conclude that the field composition of publications in Top-5 journals remains broadly stable over the period, even though citation patterns exhibit meaningful cohort trends across fields. However, we believe that a classification that relies mainly on JEL codes makes it difficult to fully characterize how economic fields and topics evolve over time—indeed, JEL codes can be used inconsistently across authors, are subject to strategic misclassification, and fail to capture emerging topics. We contribute to this literature by using a machine-learning technique to endogenously generate topics and deriving new facts about economic research. Finally, Harris (2026) maps a subset of economics science, using 24 journals and NBER working papers and unsupervised machine learning, to identify semantic relationships and flows of knowledge between topics. While this analysis is closely related to ours, our wider set of economics articles across time and journals allows us to find more topics, to characterize them more finely and find different regularities in their trajectories.

Our analysis can be extended in several directions. A first obvious extension would be to reproduce it for other fields of research than economics. The algorithmic methods we employ are not specific to the economic literature, and they could be applied as such to other fields of research, starting with physics, chemistry, biology, or mathematics, as well as to patents. A second extension would be to analyze the

extent to which key events affect the dynamics of topics. It would be particularly interesting to look at the effect of the Global Financial Crisis on the dynamics of macro-related topics, or of the recent tariff policy shift on the dynamics of trade-related topics. Extensions could also study how technological revolutions shape research topics, particularly through the growth in computing power and the increased availability of new datasets, and further distinguish between methodological topics and subject topics which may be affected in different ways. These and other extensions of our analysis are left for future research.

References

Aghion, Philippe and Peter Howitt, "Research and development in the growth process," *Journal of Economic Growth*, 1996, *1* (1), 49–73.

_ , Antonin Bergeaud, Timo Boppart, and Jean-Félix Brouillette, "Resetting the Innovation Clock: Endogenous Growth through Technological Turnover," Technical Report, Mimeo 2025.

Alperin, Juan Pablo, Jason Portenoy, Kyle Demes, and Vincent Larivière, "An analysis of the suitability of OpenAlex for bibliometric analyses," 2024.

Ando, Yoshiki, James Bessen, and Xiupeng Wang, "The Rising Returns to R&D: Ideas Are Not Getting Harder to Find," 2026. Manuscript (Journal of Political Economy submission), January 2026.

Angrist, Josh, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu, "Inside Job or Deep Impact? Extramural Citations and the Influence of Economic Scholarship," *Journal of Economic Literature*, March 2020, *58* (1), 3–52.

Arts, Sam, Nicola Melluso, and Reinhilde Veugelers, "Beyond citations: Measuring novel scientific ideas and their impact in publication text," *Review of Economics and Statistics*, 2025, pp. 1–33.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, 2003, *3* (Jan), 993–1022.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb, "Are ideas getting harder to find?," *American Economic Review*, 2020, *110* (4), 1104–1144.

_ , Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

Card, David and Stefano DellaVigna, "Nine Facts about Top Journals in Economics," Technical Report w18665, National Bureau of Economic Research, Cambridge, MA January 2013.

Chu, Johan S. G. and James A. Evans, "Slowed Canonical Progress in Large Fields of Science," *Proceedings of the National Academy of Sciences*, 2021, *118* (41), e2021636118.

Cohan, Arman, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S Weld, "Specter: Document-level representation learning using citation-informed transformers," *arXiv preprint arXiv:2004.07180*, 2020.

Culbert, Jack H., Anne Hobert, Najko Jahn, Nick Haupka, Marion Schmidt, Paul Donner, and Philipp Mayr, "Reference coverage analysis of OpenAlex compared to Web of Science and Scopus," *Scientometrics*, April 2025, *130* (4), 2475–2492.

Dieng, Adji B., Francisco J. R. Ruiz, and David M. Blei, "Topic Modeling in Embedding Spaces," *Transactions of the Association for Computational Linguistics*, December 2020, *8*, 439–453.

Dinopoulos, Elias and Peter Thompson, "Schumpeterian Growth without Scale Effects," *Journal of Economic Growth*, 1998, *3* (4), 313–335.

Fan, Angela, Finale Doshi-Velez, and Luke Miratrix, "Prior matters: simple and general methods for evaluating and improving topic quality in topic modeling," October 2017. arXiv:1701.03227 [cs].

Fort, Teresa C., Nathan Goldschlag, Jack Liang, Peter K. Schott, and Nikolas Zolas, "Growth Is Getting Harder to Find, Not Ideas," CES Working Paper CES 25-21, Center for Economic Studies, U.S. Census Bureau April 2025.

Foster, Jacob G., Andrey Rzhetsky, and James A. Evans, "Tradition and Innovation in Scientists' Research Strategies," *American Sociological Review*, 2015, *80* (5), 875–908.

Gordon, Robert, *The rise and fall of American growth: The US standard of living since the civil war*, Princeton university press, 2017.

Gordon, Robert J, "Does the "new economy" measure up to the great inventions of the past?," *Journal of economic perspectives*, 2000, *14* (4), 49–74.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, "The NBER patent citation data file: Lessons, insights and methodological tools," 2001.

Harris, Tom, "The Shape of Economics: Semantic Structure, Knowledge Flows, and Specialization," 2026. Working Paper.

Jones, Benjamin F, "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?," *REVIEW OF ECONOMIC STUDIES*, 2009.

Jones, Charles I., "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 1995, *103* (4), 759–784.

Klüppel, Leonardo and Anne Marie Knott, "Are Ideas Being Fished Out?," *Research Policy*, 2023, *52* (2), 104665.

Krauss, Alexander, "How nobel-prize breakthroughs in economics emerge and the field's influential empirical methods," *Journal of Economic Behavior & Organization*, May 2024, *221*, 657–674.

Li, Wendy C. Y. and Bronwyn H. Hall, "Depreciation of Business R&D Capital," *Review of Income and Wealth*, 2020, *66* (1), 161–180.

Lucas, Robert E, "On the mechanics of economic development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

Lucking, Brian, Nicholas Bloom, and John Van Reenen, "Have R&D Spillovers Declined in the 21st Century?," *Fiscal Studies*, 2019, *40* (4), 561–590.

Mokyr, Joel, "Secular stagnation? Not in your life," *Secular stagnation: facts, causes and cures*, 2014, *83*.

_ , "The past and the future of innovation: Some lessons from economic history," *Explorations in Economic History*, 2018, *69*, 13–26.

Park, Michael, Erin Leahey, and Russell J. Funk, "Papers and Patents Are Becoming Less Disruptive over Time," *Nature*, 2023, *613* (7942), 138–144.

Priem, Jason, Heather Piwowar, and Richard Orr, "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," *arXiv preprint arXiv:2205.01833*, 2022.

Segerstrom, Paul S., "Endogenous Growth without Scale Effects," *American Economic Review*, 1998, *88* (5), 1290–1310.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Benjamin Jones, "Atypical Combinations and Scientific Impact," *Science*, 2013, *342* (6157), 468–472.

Visser, Martijn, Nees Jan van Eck, and Ludo Waltman, "Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic," *Quantitative Science Studies*, 2021, *2* (1), 20–41. Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info. . . .

Waltman, Ludo and Nees Jan Van Eck, "A new methodology for constructing a publication-level classification system of science," *Journal of the American Society for Information Science and Technology*, 2012, *63* (12), 2378–2392.

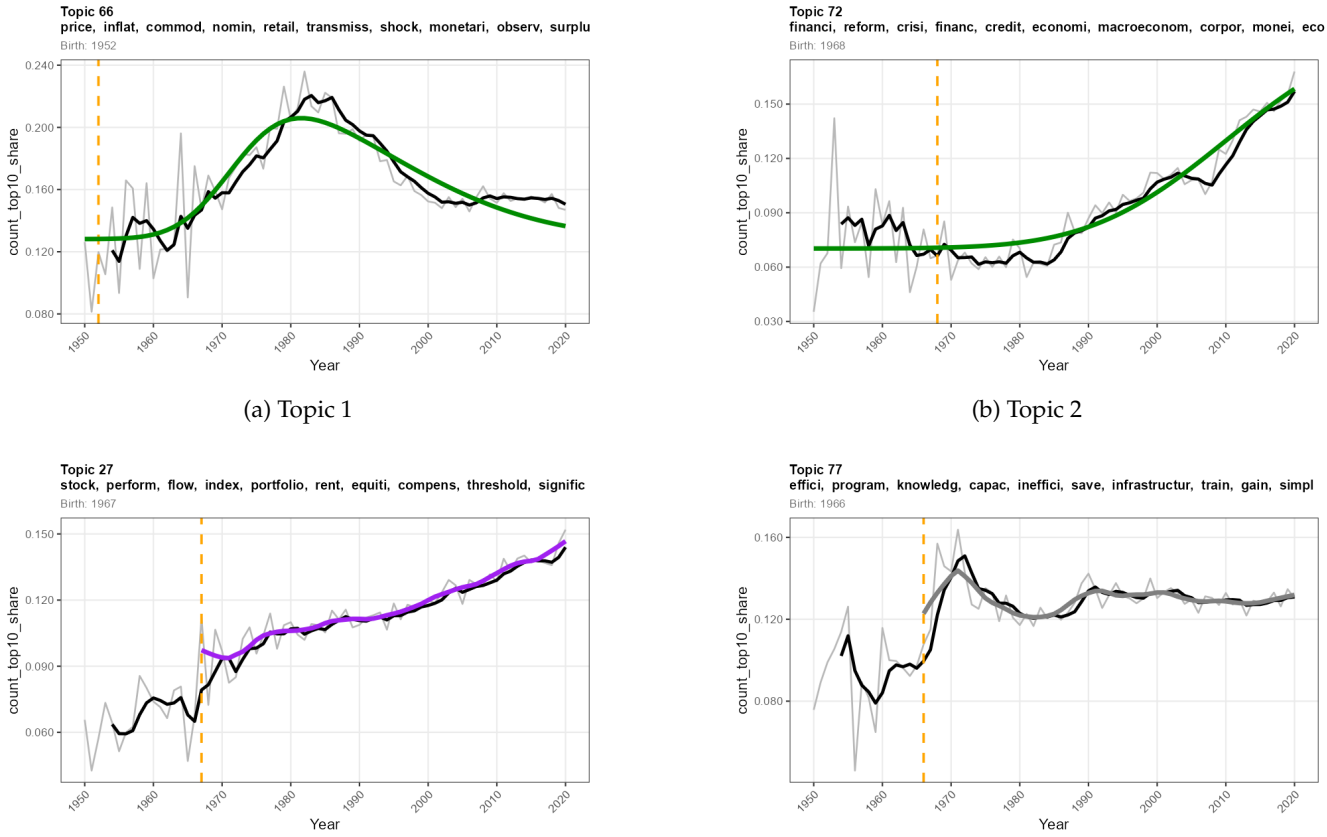
Wu, Lingfei, Dashun Wang, and James A. Evans, "Large Teams Develop and Small Teams Disrupt Science and Technology," *Nature*, 2019, *566* (7744), 378–382.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi, "The Increasing Dominance of Teams in Production of Knowledge," *Science*, 2007, *316* (5827), 1036–1039.

Zhao, Renyu and Yunxin Chen, "Accuracy Assessment of OpenAlex and Clarivate Scholar ID with an LLM-Assisted Benchmark," March 2025. arXiv:2502.11610 [cs].

Appendix

A Topic Evolution Through Years



(a) Topic 1

(b) Topic 2

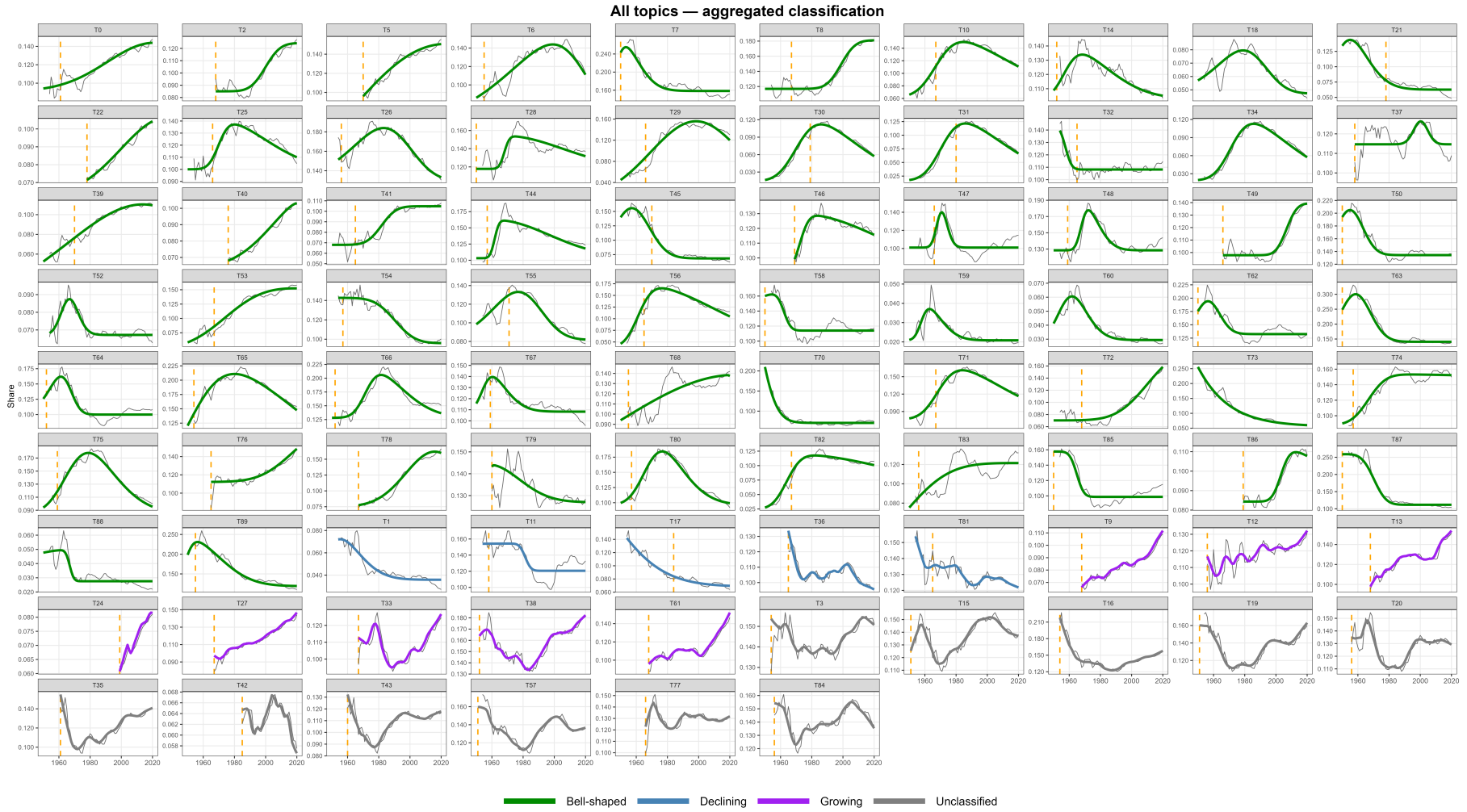
(c) Topic 3

(d) Topic 4

Figure A.1. Example of Topics over Time

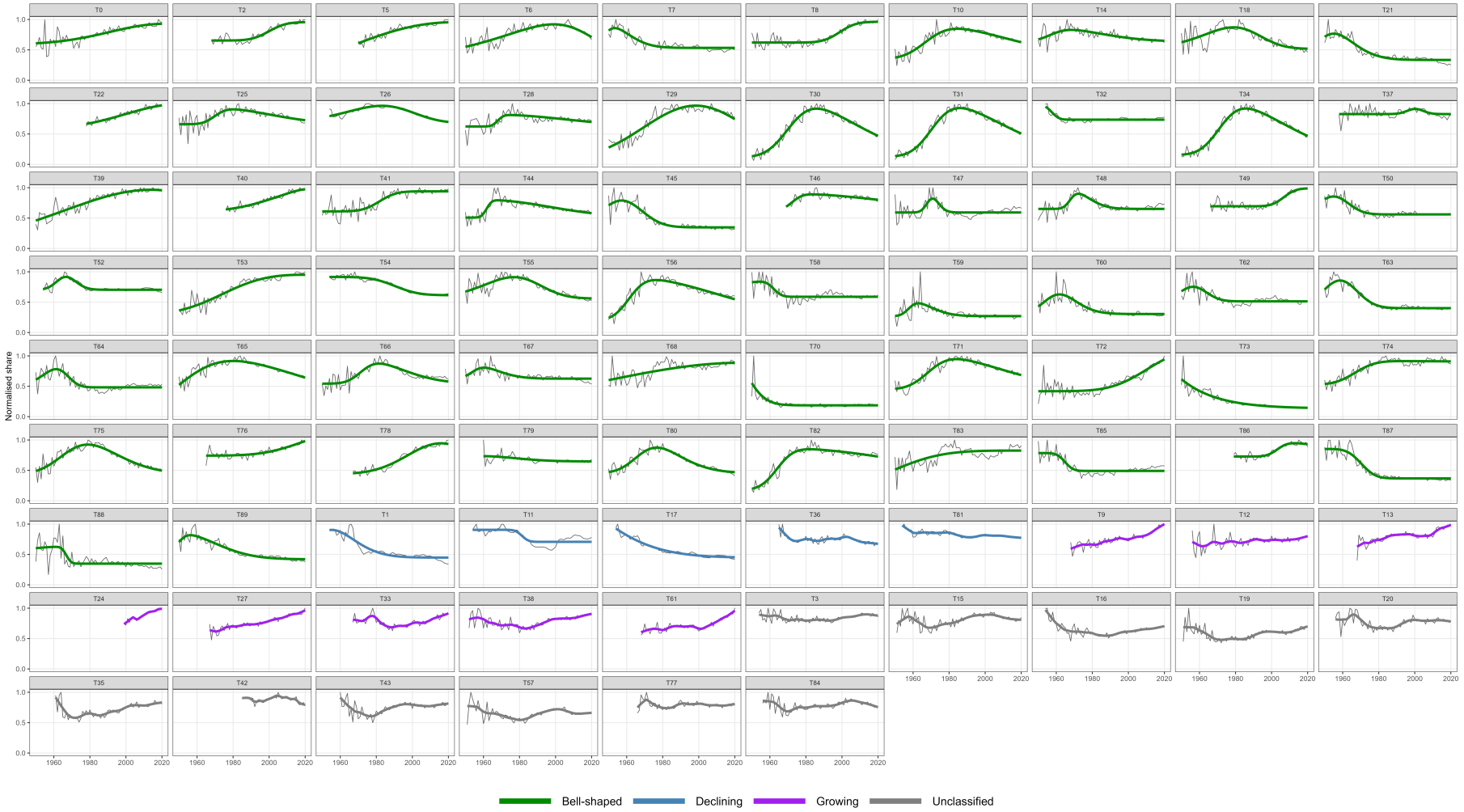
Notes: The graphs show the distribution of various topics over time, for each trajectory category that our algorithm identifies. Each graph represents the trend of a different topic, and the subcaptions provide the names of the topics.

Figure A.2. Share of Publications and Trend, All Topics



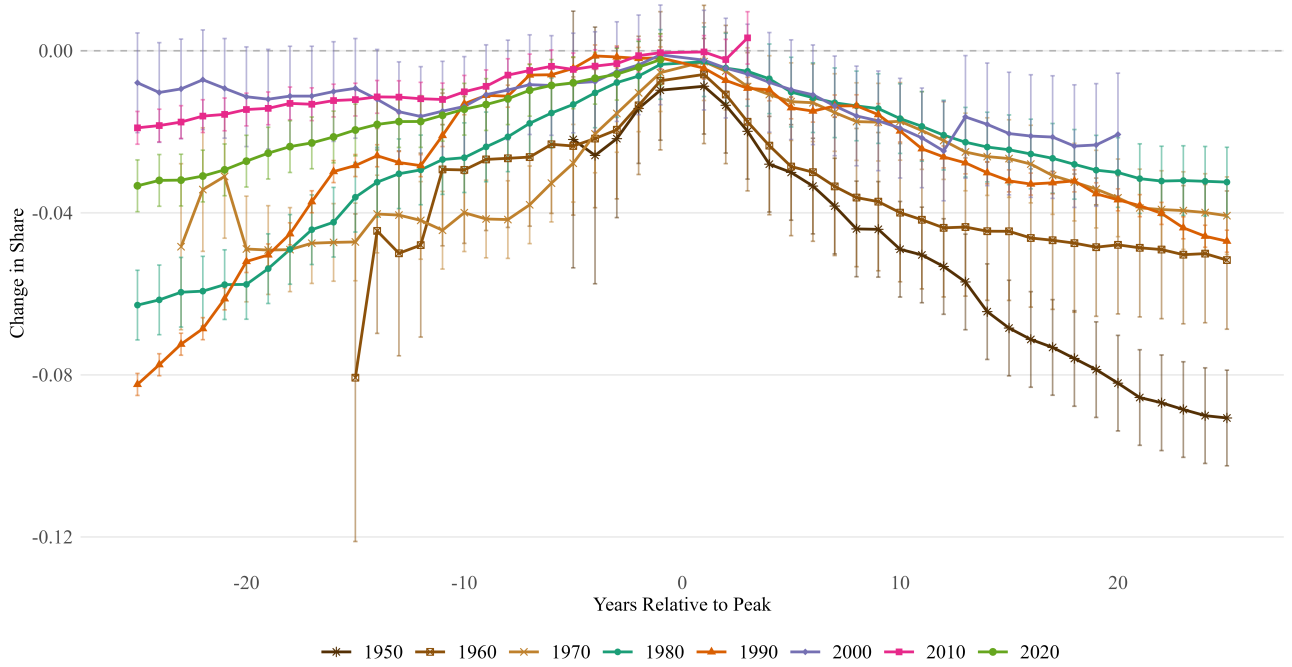
Notes: The facet graph above represent the variation of the share of publications allocated all 90 topics that we keep in our analysis. The color of the curve indicates the classification of the topic trajectory. The dotted vertical orange line represent the *birth year* of the topic as we define it in section 3.

Figure A.3. Share of Publications and Trend, All Topics, Normalized Shares



Notes: The facet graph above represent the variation of the share of publications allocated all 90 topics that we keep in our main analysis. The color of the curve indicates the classification of the topic trajectory. The dotted vertical orange line represent the *birth year* of the topic as we define it in section 3. This version displays normalized shares.

Figure A.4. Publication Share Trajectory Relative to Maximum, by Peak Decade



Notes: Each curve reports the estimated coefficients from a separate event-study regression of the smoothed top-10 publication share on relative-year indicators, estimated for each peak-decade cohort independently. Formally, for each decade d, we estimate $y_{it} = \sum_{\tau \neq 0} \beta_\tau^d \cdot \mathbf{1}[t - t_i^* = \tau] + \alpha_i + \varepsilon_{it}$, where t_i^* denotes the peak year of topic i and α_i is a topic fixed effect. The reference year is $\tau = 0$, so all coefficients measure the deviation of the publication share from its peak-year level. Topics are assigned to a decade cohort according to the calendar decade in which their publication maximum occurs. Vertical bars indicate 95% confidence intervals. The sample is restricted in that case to the topics we identify as following a bell-shaped trend. The estimation window spans 25 years on either side of the peak.

B Topic Weights and Classification

B.1 Defining a Classification for Topics

We classify topics manually into two broad categories from the topic keywords, labels, and descriptions. Method topics are those that primarily capture the tools used in the analysis, such as measurement, estimation, modelling, or evaluation. Subject topics are those that primarily capture the object of analysis: the economic question, institution, market, population, or mechanism studied in the paper. Topics that could not be assigned cleanly to either category are excluded from the type-specific panels but remain included in the pooled statistics.

Table B.1 reports the distribution of topic weights across the ten highest-weight topics assigned to each paper. Since papers can load on several topics, the unit of observation is a paper-topic pair. The first panel pools all topics, while Panels B and C report the same statistics separately for method and subject topics.

The table shows that topic weights are highly concentrated at the top of the topic ranking. In the pooled sample, the average weight of the first-ranked topic is 0.131, compared with 0.061 for the second-ranked topic and 0.041 for the third-ranked topic. Weights decline steadily after that, with ranks 6 to 10 each receiving average weights close to or below 0.02. This suggests that the first few topics capture most of the topic mass used to characterize a paper, while lower-ranked topics provide more marginal information.

The comparison between method and subject topics shows that subject topics tend to receive slightly larger weights than method topics. Across all ranks, the average weight is 0.040 for subject topics and 0.028 for method topics. This difference is also visible among first-ranked topics, although it is smaller there. This pattern is consistent with papers being primarily organized around substantive economic questions, while methodological topics often appear as secondary components of the paper's topic profile.

Table B.1. Distribution of Topic Weights by Rank and Topic Type

Topic weight	N	Mean	P25	Median	P75	SD
Panel A. All topics						
All ranks	3,652,110	0.037	0.016	0.022	0.039	0.045
Rank 1	365,211	0.131	0.076	0.108	0.159	0.085
Rank 2	365,211	0.061	0.043	0.057	0.075	0.027
Rank 3	365,211	0.041	0.030	0.039	0.049	0.016
Rank 4	365,211	0.031	0.023	0.030	0.037	0.011
Rank 5	365,211	0.025	0.019	0.024	0.029	0.008
Rank 6	365,211	0.021	0.017	0.020	0.024	0.006
Rank 7	365,211	0.018	0.015	0.018	0.021	0.005
Rank 8	365,211	0.016	0.014	0.016	0.019	0.004
Rank 9	365,211	0.015	0.013	0.015	0.017	0.004
Rank 10	365,211	0.014	0.012	0.014	0.016	0.003
Panel B. Method topics						
All ranks	1,021,133	0.028	0.014	0.018	0.028	0.033
Rank 1	58,071	0.121	0.068	0.100	0.149	0.079
Rank 2	69,932	0.057	0.039	0.052	0.070	0.027
Rank 3	79,862	0.038	0.027	0.036	0.046	0.015
Rank 4	90,092	0.028	0.021	0.027	0.034	0.010
Rank 5	98,692	0.023	0.018	0.022	0.027	0.007
Rank 6	106,916	0.020	0.016	0.019	0.023	0.006
Rank 7	115,657	0.017	0.014	0.017	0.020	0.005
Rank 8	124,618	0.016	0.013	0.016	0.018	0.004
Rank 9	133,247	0.015	0.012	0.014	0.017	0.003
Rank 10	144,046	0.013	0.011	0.013	0.016	0.003
Panel C. Subject topics						
All ranks	2,590,677	0.040	0.017	0.024	0.044	0.045
Rank 1	301,375	0.129	0.077	0.109	0.159	0.076
Rank 2	292,140	0.062	0.044	0.058	0.077	0.028
Rank 3	281,977	0.042	0.031	0.040	0.050	0.016
Rank 4	271,498	0.031	0.024	0.030	0.038	0.011
Rank 5	262,620	0.025	0.020	0.025	0.030	0.008
Rank 6	254,123	0.021	0.017	0.021	0.025	0.006
Rank 7	245,166	0.019	0.015	0.018	0.022	0.005
Rank 8	236,327	0.017	0.014	0.017	0.019	0.004
Rank 9	227,910	0.015	0.013	0.015	0.018	0.004
Rank 10	217,541	0.014	0.012	0.014	0.017	0.003

Notes: The table reports summary statistics for the weights assigned to the ten highest-weight topics in each paper. A unit of observation is a paper-topic. Rank 1 denotes the highest-weight topic in a paper, rank 2 the second highest-weight topic, and so on. Panels B and C split paper-topic weights using a Method/Subject classification.

C Topic Popularity Trajectories

C.1 Defining a Classification for Trajectories

Each of the 90 topics is assigned to one of 5 potential trajectory categories: *bell-shaped*, *declining*, *growing*, *stable*, or *unclassified*. Classification is performed using an algorithm that fits a Local Estimated Scatterplot Smoothing (LOESS)-smoothed curve to each topic's annual share of top-10 topic and tests the resulting shape against a set of parametric criteria. To maximize coverage while preserving precision, topics are processed in three sequential rounds, each applied only to topics that remained unclassified in the previous round.

Round 1 — Raw series, bell-shaped patterns detection. The first round operates on the raw, unsmoothed series of topic popularity, measured as the share of papers whose 10 most important topics comprise the topic, and targets the most unambiguous trajectory shapes: clean bell curves and flat stable trends. Trend detection and growth detection are deliberately disabled, so that only topics that have an immediately recognisable shape using the raw data are classified at this stage. A relatively strict prominence threshold of 0.15 (difference between the maximum and minimum values of the curve) ensures that only well-defined peaks are accepted.

Round 2 — Pre-smoothed series, not-bell-shaped trend detection enabled. Topics that were not classified in Round 1 are re-examined using a backward-weighted moving-average smoother (the same popularity variable, smoothed over a five-year window with differential weights, respectively $(0.1, 0.1, 0.2, 0.3, 0.3)$). Pre-smoothing attenuates year-to-year noise that may have prevented a clear shape from being identified on the raw series. Trend detection is now enabled, algorithm to flag topics whose share is monotonically declining over the observed window as *declining*. All other shape criteria—bandwidth, tolerance, prominence, and stability thresholds—are kept identical to Round 1, so that any additional classifications at this stage are attributable exclusively to the smoothing and the newly activated trend check.

Round 3 — Full detection, permissive thresholds, post-birth window. Remaining unclassified topics undergo a final pass on the raw series, now restricted to years at or after each topic's estimated birth year (the first year with at least five consecutive years of non-zero publication activity). This filter removes the pre-birth noise that can distort shape tests for recently emerged topics. Round 3 also activates growth detection enabling the identification of topics whose share is still monotonically increasing. Prominence is relaxed (`thprom = 0.08`) to capture topics with lower but genuine peaks, and the growth threshold

is lowered from 0.70 to 0.60 , making the growing category more accessible. The stability criterion is also tightened in range while the slope tolerance is substantially raised, together targeting series that are locally flat but may exhibit a slight long-run drift. Topics that remain unclassified after Round 3 are considered multimodal.

Table C.1. Parameter Settings Across the 3 Classification Rounds

Group	Parameter	Round 1	Round 2	Round 3
Input series	Variable	raw	**smoothed**	**raw (post-birth)**
	Smoothing window	none	**5-yr BWA**	none
LOESS fit	`bw` (bandwidth)	0.20	0.20	0.20
	`tol` (bell tolerance)	0.30	0.30	0.30
Peak detection	`edge` (boundary guard, yr)	0	0	0
	`thrmon` (secondary peak ratio)	0.30	0.30	0.30
	`thprom` (min. prominence)	0.15	0.15	**0.08**
Declining trend	`show_trend`	FALSE	**TRUE**	TRUE
	`trend_edge` (yr)	5	5	5
	`trend_thr_down`	0.60	0.60	0.60
Growing trend	`show_growth`	FALSE	FALSE	**TRUE**
	`grow_edge` (yr)	5	5	5
	`grow_thr_up`	0.70	0.70	**0.60**
Stable trend	`show_stable`	TRUE	TRUE	TRUE
	`stable_range`	0.10	0.10	**0.05**
	`stable_slope`	0.002	0.002	**0.50**

Notes. **bw**: fraction of the time series used as the local neighbourhood in the LOESS smoother; larger values produce smoother curves. **tol**: a topic is classified as bell-shaped only if $\mathrm{RSS}_{\mathrm{bell}} \leq (1 + \texttt{tol})\,\mathrm{RSS}_{\mathrm{cubic}}$; higher values allow noisier bells. **edge**: peaks located within `edge` years of the series boundary are excluded; `edge` = 0 disables the guard. **thrmon**: secondary peaks with prominence exceeding `thrmon` times that of the main peak trigger the multi-modal flag. **thprom**: minimum normalized prominence required to register a peak; lower values recover weaker but genuine peaks. **trend_edge** / **grow_edge**: number of years at the start of the series used to test for a declining / growing trend. **trend_thr_down**: fraction of the initial window that must be monotonically decreasing to classify a topic as *declining*. **grow_thr_up**: fraction of the series that must be monotonically increasing to classify a topic as *growing*; loosening this threshold in Round 3 captures topics still in an active expansion phase. **stable_range**: maximum normalized range of the series (peak minus trough, divided by mean) permitted for a *stable* classification. **stable_slope**: maximum absolute slope of the OLS trend line permitted for a *stable* classification; the large value in Round 3 (0.50) targets series that are locally flat despite a mild long-run drift. BWA = backward-weighted moving average with weights $(0.1, 0.1, 0.2, 0.3, 0.3)$. A dash (—) indicates that the feature is disabled. Changes are in **bold**.

C.2 Bell-shaped Trajectories

To characterize the average trajectory of bell-shaped topics around their publication peak, we estimate a parametric skew-normal model on the full set of topic-year observations. For each topic i, the smoothed publication share series is first normalized by its own within-topic maximum, so that the peak of every topic is equal to 1 regardless of its absolute size. This normalisation ensures that the estimated shape reflects the common dynamics of rise and decline rather than cross-topic differences in scale.

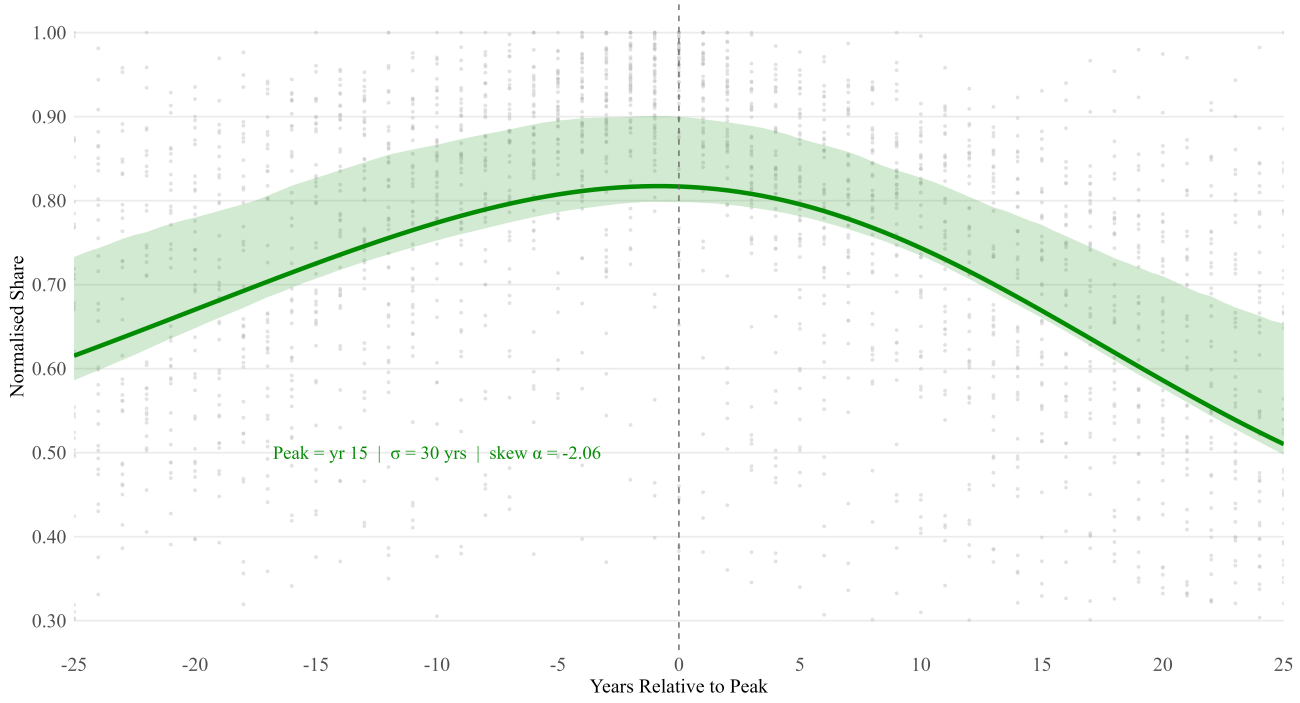
The model is then fitted jointly on all (i, t) pairs, in which t denotes years relative to the publica-

tion peak, rather than on pre-aggregated annual means. Formally, we fit the following skew-normal specification:

$$y_{it} = \frac{b_0}{\sigma} + e^{A} \cdot 2\,\phi\left(\frac{t-\mu}{\sigma}\right)\Phi\left(\alpha \cdot \frac{t-\mu}{\sigma}\right) + \varepsilon_{it} \tag{3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and cumulative distribution function, respectively. The parameter μ captures the location of the peak relative to year zero (the identified maximum), σ determines the width of the bell in years, α controls the degree and direction of asymmetry — with $\alpha < 0$ indicating a faster post-peak decline than pre-peak rise — and b_0 and A are the baseline and amplitude parameters. To ensure positivity of σ and A, both are estimated in log scale. Pointwise 95% confidence intervals for the fitted curve are obtained via parametric bootstrap.

Figure C.1. Aggregate Trend, All Topics

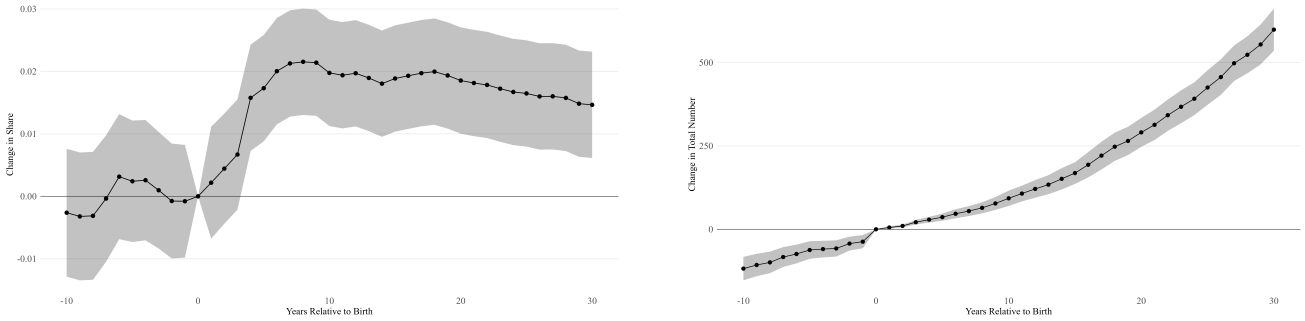


Notes: Each grey dot represents one topic-year observation, where the publication share series of each topic has been normalized by its own within-topic maximum so that the peak equals one. The green curve shows the fitted skew-normal function estimated jointly on all individual observations via non-linear least squares, with the shaded band indicating pointwise 95% confidence intervals obtained by parametric bootstrap ($B = 1{,}000$ draws). The dashed vertical line marks year zero, defined as the year in which each topic reaches its publication maximum. The estimated parameters are reported directly on the figure: μ denotes the location of the fitted peak relative to year zero, σ the half-width of the bell in years, and α the asymmetry coefficient, with negative values indicating a steeper post-peak decline than pre-peak rise.

C.3 Topics Trajectory Relative to Topic Birth

We argue in section 3 that, on average, topic popularity follows a bell-shaped curve. We also show that the magnitude of variation in percentage points is substantial (see Figure 3). Although this figure only includes topics previously identified as following a bell-shaped pattern, one might worry that the observed rise-and-fall dynamic without any subsequent increase after the peak might be a statistical artefact driven by the choice of the peak year as a reference. To address this concern and ensure that the observed magnitude is not mechanically driven by this choice, we propose an alternative reference point by identifying the birth year of each topic. We define the birth year as the first year followed by a sequence of at least 10 consecutive years during which the topic appears in at least one paper (among the 5 most likely topics). This criterion ensures that the identified birth year reflects a genuine emergence of the topic, rather than isolated or exceptional publications. Using this birth year as a reference, we compute the average trend, as shown in Figure C.2. We again observe a similar rise-and-fall pattern, where the birth year is quickly followed by a peak, and then by a slow but steady decline in the share of publications associated with the topic. We additionally reproduce Figure 5 using the birth year as reference and show that the topic trajectory is the same.

Figure C.2. Average Topic Trends, Relative to Birth Year



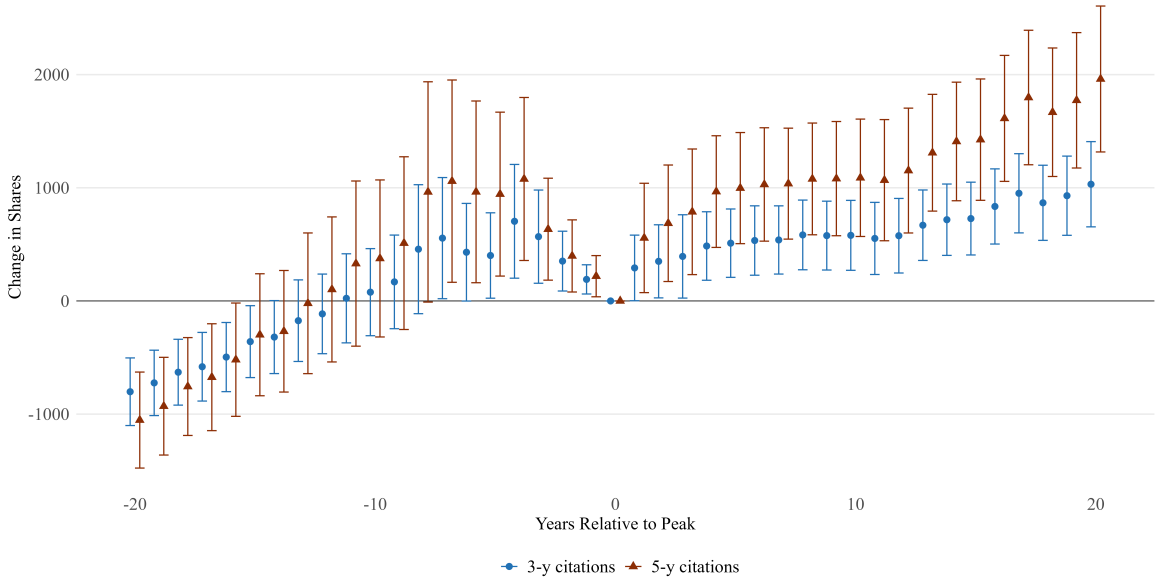
(a) Peak as Reference

(b) Birth as Reference

Notes: The graph above presents the average share of publications per topic relative to the year in which the topic is born. The share of publications is computed as the share of publications which have this topic in the 10 highest values of their distribution per year.

D Quality Evolution

Figure D.1. Average Topic Trend in Citations



Notes: The graph above presents the total number of citations per topic in the three and five years after the reference year, relative to the year in which the maximum is obtained, normalized by the topic-specific average.

E Data Collection and Treatment

In this section, we provide more information on research classification and the alignment of our sample and OpenAlex's assessment of topics.

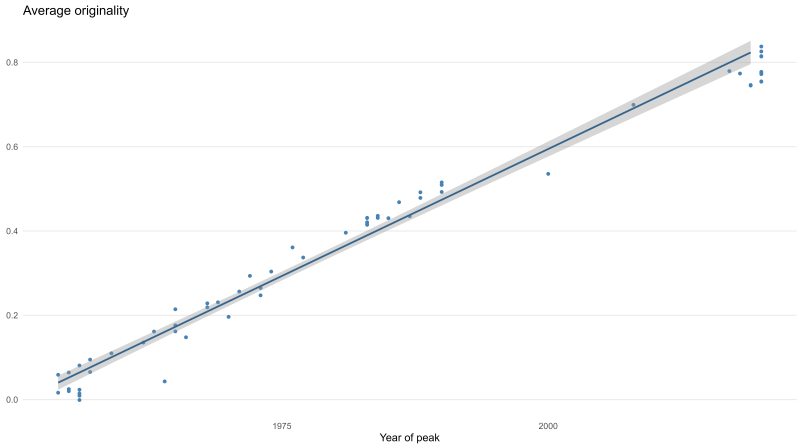
OpenAlex classifies all articles based on a pre-existing labeled dataset (Waltman and Van Eck (2012)) into a four-layered structure, the first three being the ASJC categories from Scopus, and the fourth, called "topics" (overall 107) are based on the results of a clustering algorithm performed on 10 million publications. The main reason why these topics are not suited to our analysis is that they have been trained so that the number of documents in each topic is decreasing, so that size itself is a parameter of optimization. Due to this, some topics are too small such as Economic Burden of Cancer Treatment when some are too large such Economic Development and Policy Analysis: by construction, they are not comparable. They are also computed using the citation network, which we do not want to rely on to keep topic identification based only on the text of the abstract and distinguish a topic and its inputs from other topics.

Our selection of publications based on hand-labeled journals allows us to consider papers that we

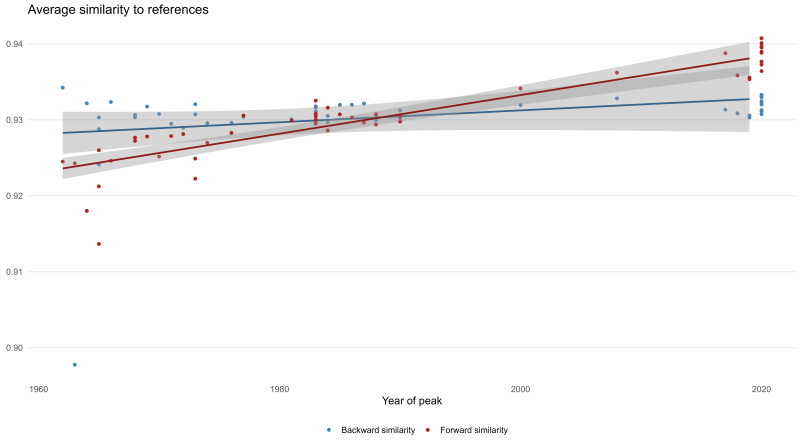
are sure belong to the discipline of economics. Figure E.1 shows that only about half of them are labeled as such in OpenAlex. As mentioned above, this also does not restrict us to authors publishing only in economics. The distribution of publications across OpenAlex topics in OpenAlex for the authors who published in our sample is even more skewed away from economics, as shown in Figure E.2. Papers tagged as economics in OpenAlex are a smaller portion of the publications of these authors.

We also provide in Figure E.4 the same graph as pane (b) of Figure E.3 and Figure E.2 for authors who have at least five publications in our sample. Yet again, although less dramatically, the number of papers outside the sample and outside of the OpenAlex Economics field is much higher than the number of papers in sample and labeled as Economics.

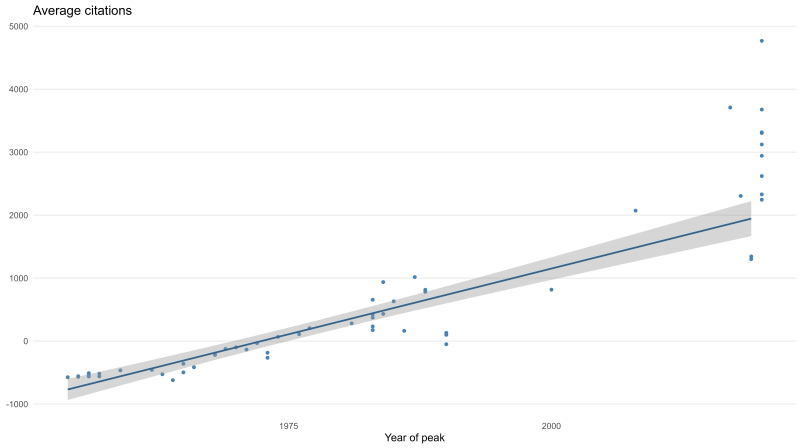
Figure D.2. Average Topic Trend: Originality and Similarity



(a) Average Topic FE in Originality



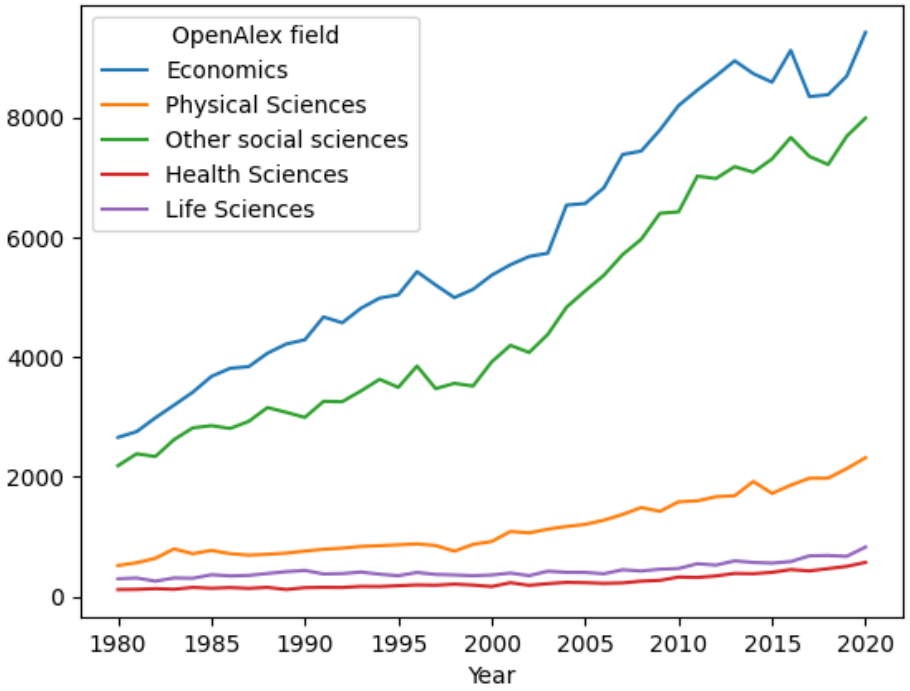
(b) Average Topic Trend in Similarity to Backward and Forward references



(c) Average Topic FE in 3-years citations

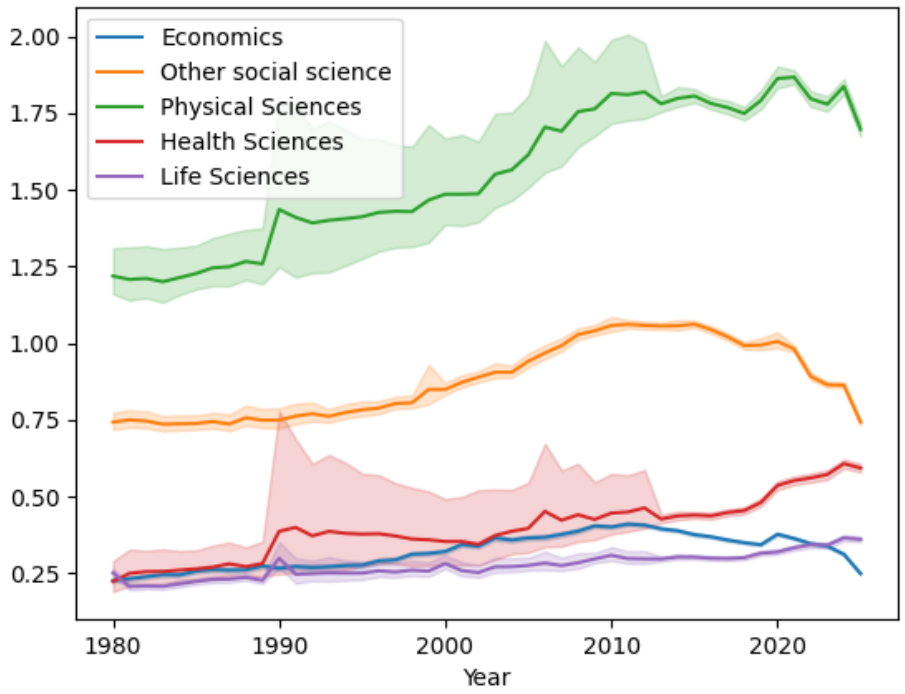
Notes: The graph above presents the topic average in average originality, similarity to references and citing papers per topic, and total citations received in the past 5 years relative to the year in which the maximum is obtained, normalized by the topic-specific average.

Figure E.1. Number of Total Publications in Sample, Per Domain



Notes: The graph above presents the total number of articles in our sample, split by whether their primary field in OpenAlex is Economics, or if not, by their domain. The shaded area represents the standard error per year.

Figure E.2. Average Number of Publications per Year per Researchers, by Domain

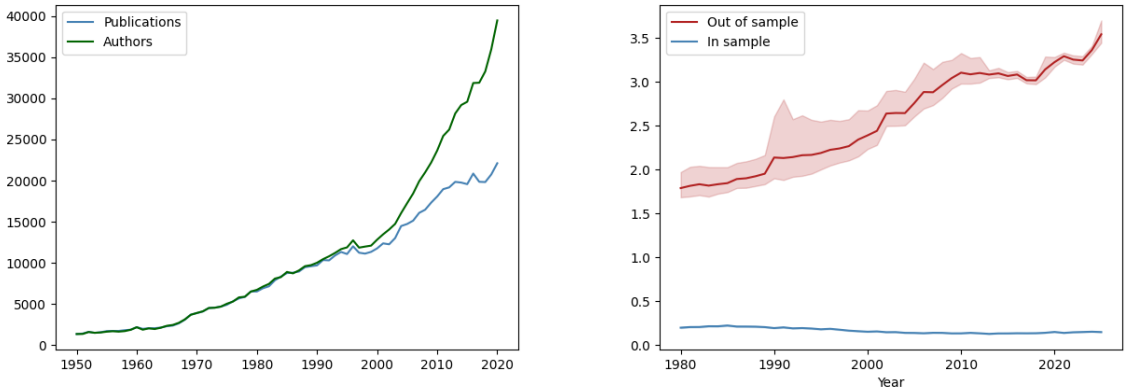


Notes: The graph above presents the average number of articles in OpenAlex per author among the ones who have published in our sample of journals, split by whether their primary field in OpenAlex is Economics, or if not, by their domain.

Figure E.3. Total number of entries in sample and in entire database

(a) In sample

(b) In the entire database

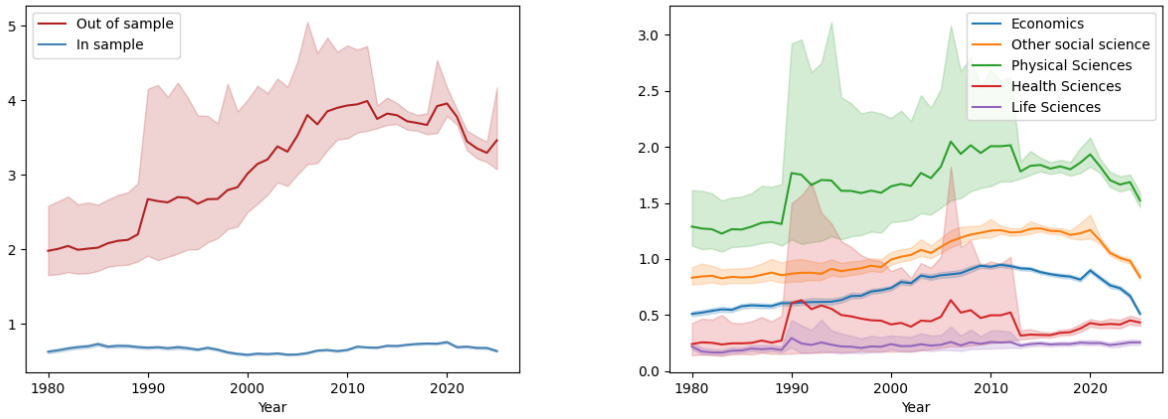


Notes: The graph above presents on the left the total number of articles (in blue) and the number of researchers who publish (in green) in our sample of journals. On the right, we plot the average number of publication per author (weighted by the number of co-authors on each publication) each year in the entire database for researchers who publish in our sample of journals, split by whether the publication is in the sample or not. The shaded area represents the standard error per year.

Figure E.4. Total Publications Count, by Sample

(a) In sample

(b) In the entire database



Notes: The graph above presents on the left the total number of articles in sample and out of sample in OpenAlex for authors publishing at least 5 times in our sample of journals. On the right, we plot the average number of articles in OpenAlex per author among the same authors who have published at least five times in our sample of journals, split by whether their primary field in OpenAlex is Economics, or if not, by their domain. The shaded area represents the standard error per year.

F Details on Topic Identification

F.1 Latent Dirichlet Allocation

Topic learning is an unsupervised machine learning process. Nevertheless, our methods make parametric assumptions about the data-generating process. The first type of topic modeling we use is Latent Dirichlet Allocation (Blei et al. (2003)). In this model, it is assumed that there are K underlying topics. Each document i draws a proportion of topics from a Dirichlet distribution, denoted θ_i. Then, for each word w_{ij} in document i at position j, one of the K topics is assigned randomly conditional on θ_i. Finally, a word is randomly drawn from the distribution of all possible words, conditional on the topic k assigned to it.

The LDA algorithm then estimates the topic distributions θ and the word distributions per topic by maximizing the marginal log-likelihood of observing the keyword distribution in the data.[25]

The logic behind LDA is not inherently semantic: it is a likelihood maximization problem. Because LDA is a **decomposition** of a text based on the contextual meaning of words in it, the resulting "topics" represent shared semantic dimensions across documents rather than their subject in a colloquial sense. For example, if many papers in a corpus contain expressions related to reporting results ("This paper finds...", "Our results show..."), that stylistic pattern could emerge as a topic—even though it is present in most documents. Such topics are not useful for our analysis, so we preprocess the data to remove uninformative dimensions. This includes removing: - Common stopwords (e.g., pronouns, "is", "the", "of"), - Rare words (those that appear too infrequently), and - Ubiquitous words (those that appear in nearly all documents).

F.2 Embedded Topic Model

The Embedded Topic Model (Dieng et al. (2020)) builds on LDA with several key differences: First, topics are not only distributions over words but also embeddings representing semantic position in the same space as word embeddings. Topic proportions are drawn from a log-normal distribution instead of a Dirichlet. Then, the word distribution per topic is computed based on the dot product between the word embedding and the topic embedding—i.e., words that are semantically closer to the topic in embedding space are more likely to occur.

[25] The algorithm maximizes the marginal log-likelihood because the "true" likelihood depends on latent variables (i.e., topic assignments), which are unobservable. Furthermore, since the marginal log-likelihood is intractable, the algorithm maximizes the Evidence Lower Bound (ELBO), which is equivalent in effect.

The algorithm then estimates the topic embeddings, topic proportions per document, and word distributions per topic to maximize the marginal log-likelihood of the data.[26]

F.3 Algorithmic performance metrics

F.3.1 Log-likelihood

The most obvious metric for the evaluation of a model is its target; here, it's the maximum marginal log-likelihood obtained for a model K in this dataset. In LDA and ETM, the log-likelihood should be higher if the number of topics is the "true one"; if the data generating process is exactly as specified, the global maximum in log-likelihood should be at the point K that represents the true number of topics. In both models, the marginal log-likelihood is given by:

$$\mathcal{L}(\alpha, \rho) = \sum^{D} \log p(\mathrm{w}_d | \alpha, \rho)$$

In which α and ρ are parameters of the model (in LDA, α is a vector of parameters of the Dirichlet distribution of topics and ρ, called β in the original paper, is the matrix representing the distribution of words over topics; in ETM, α is the vector of topic embeddings and ρ is the matrix of word embeddings), D is the number of documents d and $p(\mathrm{w}_d | \alpha, \rho)$ is the probability of observing the realised word distribution w_d in document d given α and ρ. While this is not what the algorithm actually maximizes for practical reasons, both algorithms allow us to retrieve this metric.

F.3.2 Perplexity

The perplexity of a model is an indicator for generalisation performance of the model: it is a measure of how good the model is at predicting topics from a similar dataset. It is also based on likelihood, as the **inverse** of the log-likelihood of the model computed by the algorithm on a set of documents that it was not trained on. If the data generating process of the model also applies to the out-of-sample documents, then the model is generalizable, so that log-likelihood will be high and perplexity will be low. For a model D to be evaluated, on a set of M of documents d with N_d words distributed according to w_d, Blei et al. (2003) define perplexity as :

[26] *Idem.*

$$perplexity(D) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(\mathrm{w}_d)}{\sum_{d=1}^{M} N_d}\right)$$

Usually, in LDA, perplexity keeps decreasing as the number of topics increases, but the decrease in perplexity should slow sharply after the optimal number of topics is passed.

F.4 Topic-specific validity metrics

Because our aim using these metrics is not only to get a topic distribution with a high likelihood, but also a distribution that is interpretable, we also use topic coherence. Broadly speaking, it is a measure of how much documents within a topic are alike. Several metrics are used for this purpose. Here, we will use two different coherence metrics, for LDA and for the ETM. Coherence is not explicitly maximized for both of these methods.

The first coherence metric, which we apply to LDA, is the UMass coherence metric (**?**). For each topic k, we take the M words $\{w_1^k, w_2^k, ..., w_M^k\}$ that are the most frequent in the topic (usually the top 10 words) and compute:

$$C_{UMass} = \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{i=1}^{M}\sum_{j=i+1}^{M}\log\frac{D(w_i^{(k)}, w_j^{(k)})+1}{D(w_i^{(k)})}\right)$$

In which $D(w_i^{(k)}, w_j^{(k)})$ is the number of co-occurrences of $w_i^{(k)}$ and $w_j^{(k)}$ in the same document in k, and $D(w_i^{(k)})$ is the number of occurrences of $w_i^{(k)}$ in a document in k.

Following Dieng et al. (2020), for the ETM, we use the following metric of topic coherence, usually called NPMI coherence. The notations remain the same.

$$C_{NPMI} = \frac{1}{K}\sum_{k=1}^{K}\frac{2}{P(M,2)}\sum_{i=1}^{M}\sum_{j=i+1}^{M} f(w_i^{(k)}, w_j^{(k)})$$

in which K is the number of topics, $P(M,2)/2$ is the number of (unordered) permutations that can be done of two words in a set of M words, and $f(w_i, w_j)$ is the normalized pointwise mutual information function defined as:

$$f(w_i, w_j) = -\frac{\log\frac{P(w_i,w_j)}{P(w_i)P(w_j)}}{\log P(w_i, w_j)}$$

In which $P(w_i, w_j)$ is the probability of words w_i and w_j of occurring together (co-occurrence probability), and $P(w_i)$ is the marginal probability that word w_i will appear.

G Optimization of the Embedded Topic Model

G.1 Optimization

In practice, the embedded topic model to be estimated on a dataset of documents takes three parameters: the number of topics K to find, the minimum document frequency and the maximum document frequency of words to be included in the dataset – i.e. we remove words that appear in less than a given share of documents and in above a given share of documents. As mentioned above, we iterate in order to find the specification that fits our data best. We iterate over the combination of: numbers of topics from 20 to 300, by steps of 5; minimum document frequencies of 0.01, 0.05, 0.1, 0.5, and 1%, and maximum document frequencies of 25, 50 and 80%. This means that we have to compute 855 different models in order to assess which is the best.

Due to the fact that this optimization process becomes longer the more topics to be computed and the larger the vocabulary to be taken into account, we randomly select a subset of the data prior to applying the algorithm (20% of the data per year i.e. 70,632 abstracts) and split it into a train dataset and test dataset which are the same for all iterations of the algorithm. The number of training epochs is also set at the same level of 100 for all iterations. There is no validation dataset as the process is unsupervised. The full process of iterations with the module embedded_topic_model[27] takes about 14 days, using SPECTER 2 embeddings.

Figure G.1 shows the results of the algorithm, summarising the efficiency of the model by presenting the ratio of coherence to perplexity. Our final model is also shown on the graph. It is clearly close to the maximum of the ratio, but not exactly at that point. We now explain our method to choose the optimal model.

For a fixed maximum document frequency, Figure G.2 shows that up until we reach the 1% minimum document frequency level, coherence appears to improve , with only marginal returns for the 0.5% levels. A similar pattern is occurring with perplexity, which improves for higher values of minimum document frequency, but much less marginally. Therefore, we set our optimal minimum document frequency at 1%.

Maximum document frequency, after the cleaning and removal of stopwords, does not change the

[27] We use the Python package developed by Luis Mato, with minor adjustments allowing us to keep the same train/test split for every algorithm.

Figure G.1. Coherence to perplexity ratio for all specifications

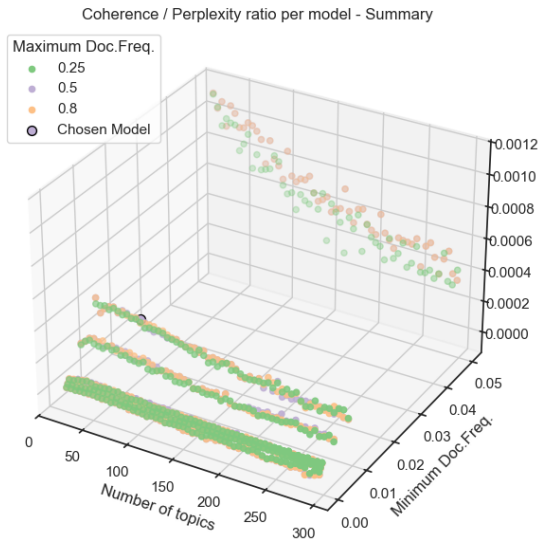
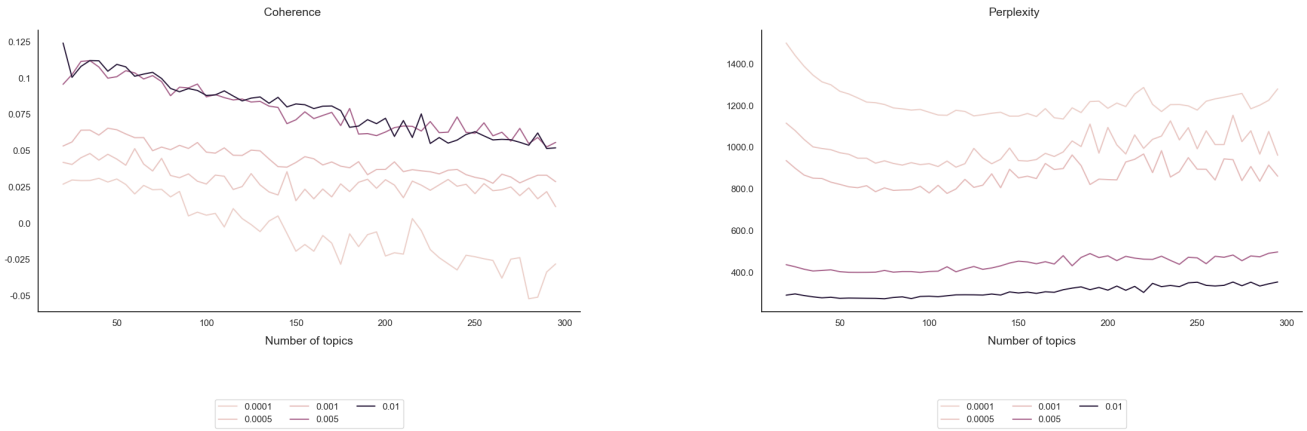


Figure G.2. Coherence and Perplexity Analysis for Each Model, by Minimum Frequency



(a) Coherence

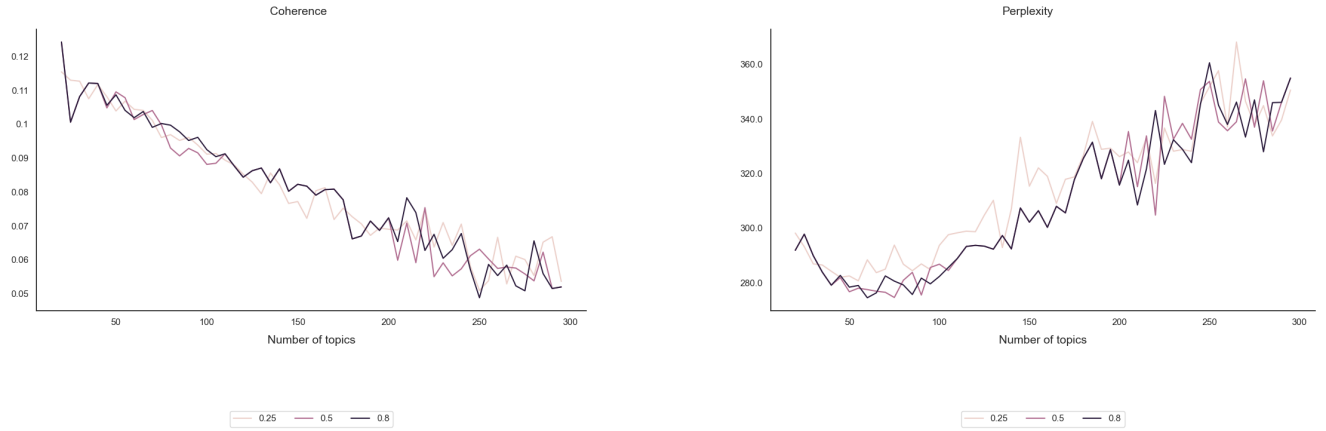
(b) Perplexity

Notes: This figure displays the relationship between coherence and perplexity for each model as a function of the number of topics. The analysis is based on a minimum document frequency (Min. DF) threshold varying from 0.01% to 1% and a maximum document frequency (Max. DF) of 50%.

vocabulary taken into account by the model very much. Figure G.3 shows the two metrics as a function of maximal document frequency and number of topics, for models at our optimal minimum document frequency of 1%. While the difference is not huge, the 50% models appear to perform slightly better over the different numbers of topics. We set the maximal document frequency at 50%

All of this being set, we now look at which number of topics may be optimal. The right pane of Figure 2 shows that the minimum value for perplexity is obtained around 75, and a secondary local

Figure G.3. Coherence and Perplexity Analysis for Each Model, by Maximum Frequency



(a) Coherence

(b) Perplexity

Notes: This figure displays the relationship between coherence and perplexity for each model as a function of the number of topics. The analysis is based on a minimum document frequency (Min. DF) threshold of 1%.

minimum appears at 90 topics, before monotonically rising after. Therefore, we select 90 as the optimal number of topics in our model.

G.2 Robustness and Validation on Simulated Data

In order to assess the effectiveness of our optimization process in choosing the right number of topics for a dataset, we perform the entire algorithm on two a dataset generated using ChatGPT. We generate 3,000 abstracts for each of the three following subjects: general macroeconomic growth inspired by 1950s research, the rational expectations literature from the 1970s, and the credibility revolution and RCTs literature from the 2000s.

We make ChatGPT write a set of 12,000 abstracts equally split in three subjects. While we know for sure that all of those abstracts are in theory meaningful, we still apply the filtering process described in section 2.1, which leaves us with 11846 publications. This first suggests that only a very small number of abstracts are wrongly excluded from the dataset by our filtering process. We then apply the topic clustering algorithm from subsection 2.2. Figure G.4 displays the results of the algorithm. There is clearly a peak at 3 topics for every combination of minimum and maximum document frequency.

We use this dataset in order to check whether the process would detect inverted-U shaped topic dynamics when they are specified in the data-generating process. In order to have such shapes, for each of the three subjects, we randomly assign a date from Gaussian distributions centered around respectively 1955, 1985, and 2020. Using the optimal model obtained on this sample, we find that these curves are

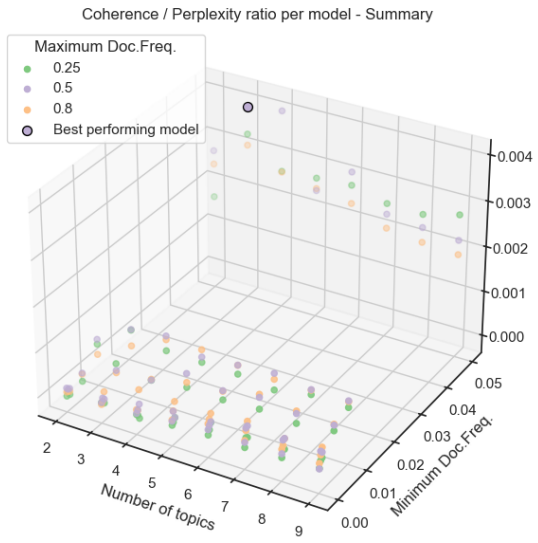
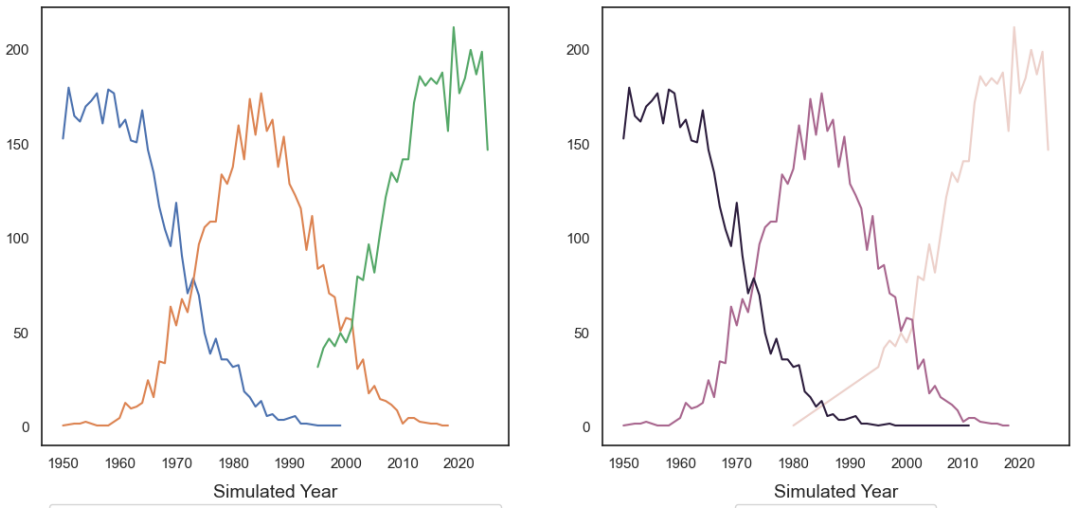


Figure G.4. Coherence to perplexity ratio for all specifications in the three-subjects simulated data

almost perfectly replicated as shown in Figure G.5.

Figure G.5. Number of abstracts per simulated subject and simulated date (left) and by algorithm subjects and simulated date (right)



We also use these abstracts for the following sampling exercise. In reality, the number of economics abstracts has risen over time. This means that topics peaking earlier may feature fewer papers than those that peak in more recent time. We want to test whether this fact biases the topic assignment algorithm towards newer topics. In order to do so, we sample the simulated abstracts with a probability increasing linearly each year from 20% in the minimal simulated year to 100% in the last simulated year. We then

apply our topic assignment algorithm to this subset of the simulated data. The algorithm performs just as well.

H Alternative measures

In this section, we present alternative measures of novelty based on textual analysis.

Our first alternative measure comes from Arts et al. (2025). It is the average semantic distance to all recorded publications in OpenAlex in 2024, computed based on embeddings from the same model as the one that we use in our classification algorithm. Figure H.1 shows that contrarily to our other measures, this measure of novelty decreases. However, this is not inconsistent with a lack of decline of average quality in economics; if the average paper in economics becomes more similar to the average paper in computer science, mathematics or sociology, this measure could increase due to an increase in generality in papers in economics at the field level.

Figure H.1. Average Semantic Distance to all OpenAlex Papers