École des Hautes Études en Sciences Sociales

Doctoral School 465 – Économie Panthéon-Sorbonne

Paris-Jourdan Sciences Économiques, UMR 8545

Ph.D.

Discipline: Economics

RAPHAËL WARGON

Essays in the economics of science

Thesis supervised by: Philippe Aghion

Date of defence: 16 December 2026

Referees
1 Bronwyn Hall, UC Berkeley
2 Reinhilde Veugelers, KU Leuven

Jury
1 Philippe Aghion, INSEAD, LSE, PSE
2 David Margolis, CNRS, Paris 1
3 Fabiana Visentin, Maastricht University

École des Hautes Études en Sciences Sociales

École doctorale 465 – Économie Panthéon-Sorbonne

Paris-Jourdan Sciences Économiques, UMR 8545

pour l'obtention du grade de docteur en sciences économiques

Discipline : Sciences économiques

RAPHAËL WARGON

Essais en économie de la science

Thèse dirigée par : Philippe Aghion

Date de soutenance : le 16 décembre 2026

Rapporteures	1	Bronwyn Hall, UC Berkeley
	2	Reinhilde Veugelers, KU Leuven

Jury	1	Philippe Aghion, INSEAD, LSE, PSE
	2	David Margolis, CNRS, Paris 1
	3	Fabiana Visentin, Maastricht University

REMERCIEMENTS

Je voudrais commencer par exprimer ma profonde reconnaissance envers mon directeur de thèse, Philippe Aghion, tout d'abord pour ses nombreux conseils, son soutien constant, ses idées bouillonnantes et la clarté de sa réflexion, sur laquelle j'ai toujours pu compter pour me mener vers un travail plus accompli et plus intéressant. J'aimerais aussi le remercier pour la générosité qu'il a montrée dans toutes les dimensions de notre relation. Pour commencer, sa générosité dans la confiance qu'il m'a accordée dès mon arrivée dans son laboratoire, en me permettant de travailler directement avec lui sur un projet dont j'ai pu devenir un co-auteur. Ce n'est en revanche pas une exception dans notre laboratoire, où le travail des jeunes est reconnu et où ils sont poussés avec tout l'enthousiasme qui caractérise Philippe. C'est à la fois son projet et cet environnement qu'il a créé qui m'ont donné goût à la recherche et qui m'ont guidé tout au long de ces quatre années et demie de thèse. Cette confiance s'est poursuivie au fur et à mesure qu'il m'a aidé à m'affirmer au sein du laboratoire. Ensuite, sa générosité dans les opportunités, en me proposant notamment de travailler sur de nombreuses idées passionnantes, parmi lesquelles celle du chapitre qui clôt cette thèse, qui m'a emballé dès la première conversation, et que j'ai hâte de mener avec lui dans d'autres directions. Enfin, je le remercie pour la générosité de son tempérament, dans sa gentillesse et dans son approbation, et pour son optimisme de combat qui compense un peu mon pessimisme personnel. J'espère que nous continuerons à collaborer longtemps encore.

Je souhaite également remercier Bronwyn Hall et Reinhilde Veugelers pour l'honneur immense qu'elles me font en acceptant d'être rapporteures de cette thèse. Chacune a contribué à la genèse de ces projets. C'est d'abord sur leurs travaux fondateurs que repose une grande partie de ma recherche. Ensuite, leurs retours à la fois détaillés et ouverts sur cette thèse m'ont permis de considérablement l'améliorer. J'adresse aussi de sincères remerciements à David Margolis, non seulement parce qu'il a accepté de présider ce jury, mais également pour les heures de conseils qu'il m'a accordées au sujet du deuxième chapitre de ce travail, et dont je suis toujours sorti avec beaucoup d'optimisme et de corrections productives. Je remercie également Fabiana Visentin pour notre conversation sur son travail et l'économie de la science en général lors de notre rencontre, et pour avoir accepté de siéger dans le jury de cette thèse.

Je souhaite aussi remercier tous les chercheurs qui m'ont accompagné au long de cette thèse. D'abord, j'adresse ma gratitude à mes co-auteurs durant mon doctorat, Céline Antonin, Antonin Bergeaud, Gaétan de Rassenfosse, David Strömberg et Xueping Sun, pour leur expertise, leur soutien et leur aide. Je remercie Clément Malgouyres, dont la bienveillance et les commentaires lors de sa participation à mon comité de thèse m'ont permis d'améliorer substantiellement ces chapitres. Je remercie aussi les chercheurs de PSE qui m'ont conseillé, notamment Francis Bloch et Catherine Bobtcheff. Merci également aux chercheurs seniors associés au laboratoire Farhi pour l'innovation et de l'INSEAD et aux participants à notre séminaire commun, Martin Aragoneses, Dylan Glover, Francis Kramarz, Jacques Mairesse, Clément Mazet-Sonilhac, Alexandra Roulet, et Dimitrije Ruzic, pour leurs commentaires tout au long de ma thèse, ainsi qu'aux chercheurs externes qui m'ont conseillé, Yuchen Lin, Kyle Myers, Fabian Waldinger et Erina Ytsma.

Je souhaiterais aussi remercier l'intégralité du laboratoire Farhi pour l'innovation. En particulier, j'adresse ma profonde gratitude à celui qui a écrit avec moi la plus grande partie de mes travaux, mon collègue, co-auteur, co-organisateur, co-conspirateur, parfois colocataire, camarade et ami Luc Paluskiewicz, dont la vivacité d'esprit, l'inventivité, la précision intransigeante, et le sens de l'humour improbable font partie des raisons principales pour lesquelles j'ai décidé de me lancer dans la thèse. Ton amitié fait de moi un meilleur chercheur et une meilleure personne et je te serai toujours reconnaissant pour ces années passées à travailler ensemble. J'adresse aussi des remerciements particuliers à Maxime, dont la camaraderie, l'écoute et les conseils m'ont sorti de l'impasse et du pessimisme un grand nombre de fois depuis ma première année au laboratoire. Merci à Simon B, Emma, Timothée, Simon M, Antoine M et Olimpia pour leur accueil, et à Florian, Louise et Souleymane pour leur aide précieuse et leur travail minutieux sur les chapitres 2 et 3 de cette thèse. Merci à Anaïs, Antoine T, Claudia, Corentin et Milla d'avoir été des compagnons de bureau parfaits supportant mes commentaires, râleries et fredonnements tous les jours, à Paul avec qui je compte bien me battre pour le projet sur l'espace, à Gabriela et Benjamin d'avoir rendu le 5^{e} étage plus animé et joyeux pendant l'année écoulée. Merci à Patrizia de mettre de l'ordre dans notre chaos et de me donner le sourire chaque jour où nous sommes ensemble dans mon nouveau bureau. Merci à Adrien, Alexis, Alice, Andrea, Antoine H, Antonio, Asmae, Aymann, Eglantine, Enrichetta, Gustave, Justine, Laure, Loris, Miguel, Monica, Raphaëlle, Romain, Ricardo, Sandra, Solomiya, Sonja et tous les jeunes du laboratoire, ainsi qu'aux camarades doctorants et doctorantes de PSE et d'ailleurs : à María José González-Fuentes pour son enthousiasme, sa créativité et sa prévenance, à Bertille Évreux pour le partage du stress de fin de thèse, des galères avec les étudiants et des bons moments en dehors, et à Lucas Pérez, Pascale Champalaune et Louis Sirugue.

Je voudrais également remercier toutes celles et ceux qui m'accompagnent en dehors de mon travail, Léonard, Abdelilah, Lyes, Joseph et Mathéo, Dr Amandine Le Bellec, Cybèle, Marion et Raphaëlle, Anne, Lise, Euryale et Tomislav, Lorine, Antoine et Nicolas, Elizabeth, Hugo et Mathilde. Merci à Brigitte et Julia pour leur accueil dans leur famille et les bons moments passés ensemble, et à Jérôme qu'on est heureux d'avoir avec nous. Merci à Nadia, pour ses encouragements au long cours. Merci à la famille Wargon de Saint-Maur, Alexis et Déborah, Samuel, Max, Joshua et Ezékiel, que j'aime tous très fort. Je pense encore à mes grands-parents, Maurice et Suzanne, et Lionel et Francine, et j'espère qu'ils seraient fiers.

Merci à ma mère pour son amour infini et inconditionnel, ses conseils sans jugement, sa douceur, son intelligence et sa finesse. Merci d'avoir été mon pilier pendant toute ma vie, d'avoir toujours eu confiance en moi, d'avoir toujours été de mon côté. Cette thèse te doit beaucoup, comme tout le reste. Merci à mon père pour son amour infini et inconditionnel, ses conseils remplis de jugement, sa franchise, son intelligence et son intégrité. Merci d'avoir toujours été exigeant, de m'avoir appris à ne pas me prendre au sérieux, et de m'avoir montré depuis tout petit qu'il fallait aller au bout de ce que l'on pense. Merci à mon petit frère Solal qui a réussi à publier en six mois alors qu'il m'a fallu cinq ans, et à qui je souhaite tous les succès imaginables. Merci à ma grande sœur Hannah qui est toujours mon modèle en tout, que j'aime, que j'adore et que j'admire. Je suis très chanceux et très heureux d'être entouré par cette famille.

Enfin, merci à mon compagnon, Quentin. Je n'aurais pas pu écrire cette thèse sans ton soutien indéfectible. Depuis sept ans, nous avons partagé toutes les joies, et dans toutes les difficultés, j'ai pu compter sur le réconfort de ta main dans la mienne. Mes journées sont plus simples et plus légères depuis que nous les commençons et terminons ensemble. Tu rends ma vie douce.

Résumé et mots clés

Titre

Essais en économie de la science

Résumé

Dans le paradigme schumpétérien, la croissance économique repose sur l'innovation, qui dépend elle-même de l'avancée de la science. Si ce paradigme rend endogène la production de recherche et développement, la production du savoir scientifique sur laquelle elle s'appuie reste largement une boîte noire. Alors que le nombre de publications a crû de façon exponentielle, plusieurs inquiétudes pèsent sur cette tendance : des gouvernements populistes s'en prennent aux scientifiques ; le resserrement des budgets publics pousse les gouvernements à réformer l'organisation de la recherche ; enfin, les idées nouvelles pourraient devenir plus difficiles à trouver à mesure que le savoir s'accumule. À partir de données bibliométriques à grande échelle, cette thèse évalue dans quelle mesure la production scientifique dépend de son environnement, en étudiant l'effet de chacun de ces trois enjeux sur la recherche académique.

Le premier chapitre étudie comment les attaques contre la liberté académique nuisent à la production des idées, à travers le cas des chercheurs hongrois après le retour au pouvoir de Viktor Orbán en 2010. En combinant des bases nationales de publications et une base bibliométrique internationale, nous montrons que les trajectoires des chercheurs divergent fortement selon leur alignement politique perçu, déduit de leurs affiliations, de leur appartenance à des groupes politiques et de leurs sujets de recherche. La quantité et la qualité des publications des chercheurs perçus comme opposants diminuent, ce que nous attribuons à l'érosion de leur réseau de coauteurs et au recul de leur influence. Ils réagissent en quittant le pays, en exprimant leur opposition dans leurs travaux ou en s'autocensurant. Comparés à des chercheurs slovènes et croates semblables, même les chercheurs hongrois qui ne sont pas perçus comme opposants publient davantage dans des revues de langue hongroise, moins bien classées, et émigrent plus souvent.

Le deuxième chapitre étudie comment l'autonomie modifie la production d'un service public, à travers la réforme des universités françaises de 2007. Celles-ci ont obtenu davantage de latitude dans leurs recrutements et, à partir de 2011, la possibilité de concourir en consortium aux Initiatives d'Excellence (IDEX). Les grandes écoles et les organismes nationaux de recherche étant exclus de la réforme, j'en identifie l'effet par des différences-en-différences échelonnées sur des données OpenAlex. L'autonomie rend les universités plus attractives : en dix ans, leur nombre de chercheurs actifs augmente de 15 % et leurs publications de 23 %, sans changement significatif des publications par chercheur ni de leur qualité. Les publications des chercheurs déjà en poste avant la réforme chutent en revanche de 40 %, et leur qualité baisse pendant près de dix ans. Les gains institutionnels relèvent donc d'effets de composition : l'autonomie augmente la taille des institutions, mais réduit la productivité des chercheurs déjà en place.

Le troisième chapitre contribue au débat sur les rendements décroissants de la production d'idées en étudiant deux disciplines, l'économie et la chimie. Par des méthodes de traitement automatique du langage, nous classons 360 000 articles d'économie et 2,5 millions d'articles de chimie publiés entre 1950 et 2020 en 90 et 80 sujets, chacun caractérisé par un ensemble cohérent de thèmes et de vocabulaire. Trois faits communs aux deux disciplines en ressortent. D'abord, les sujets suivent souvent un cycle de vie en cloche : leur part dans la discipline croît, atteint un pic, puis décline. Ensuite, même ces sujets continuent de croître en valeur absolue, sans baisse apparente de qualité. Enfin, de nouvelles combinaisons de sujets apparaissent continuellement. Ces résultats éclairent le débat sur la raréfaction des idées nouvelles.

Mots clés

Économie de la science ; Économie publique ; Bibliométrie ; Liberté académique ; Hongrie ; Autonomie des universités ; Productivité de la recherche ; Analyse textuelle

Abstract and Keywords

Title

Essays in the economics of science

Abstract

Economic growth relies on innovation, which itself relies on science. While Schumpeterian growth theory makes research and development endogenous, the production of the scientific knowledge on which it builds remains largely a black box. Although the number of scientific publications has grown exponentially over the past decades, several concerns have arisen about the sustainability of this trend: populist governments increasingly target academics; tightening public budgets lead governments to reform how research is organised; and ideas may become harder to find as knowledge accumulates. Using large-scale bibliometric data, this thesis assesses how much scientific production depends on its environment by studying how each of these three issues affects academic research.

The first chapter analyses how attacks against academic freedom harm the production of ideas, focusing on Hungary after Viktor Orbán returned to power in 2010. Combining national research repositories with international bibliometric data, we show that academics' trajectories diverge sharply depending on their perceived political alignment, inferred from their institutional affiliations, memberships of political groups and research topics. Academics perceived as political opponents experience a large decrease in the quantity and quality of their publications, which we attribute to a decline in their number of co-authors and in their scientific influence. They respond by leaving the country, by expressing dissent in their research, or by self-censoring. Compared with similar researchers in Slovenia and Croatia, even Hungarian researchers not perceived as opponents shift their publications towards lower-ranked national-language journals and are more likely to leave the country.

The second chapter studies how replacing central administration with local autonomy affects the output of a public service, through the 2007 reform of French universities. Universities gained more control over hiring and, from 2011, could compete in consortia for large institution-level grants. Since *Grandes Écoles* and national research institutions were excluded from the reform, I identify its effect with a staggered difference-in-differences design on OpenAlex data. Autonomy makes universities more attractive: within ten years, their number of researchers rises by about 15% and their total publications by 23%, while publications per researcher and quality do not change significantly. Publications by researchers already in place before the reform fall by 40%, and their quality declines for up to ten years. Institution-level gains are therefore compositional: autonomy raises total output by attracting new researchers while reducing the output of incumbents.

The third chapter contributes to the debate on the production of ideas by examining two disciplines, economics and chemistry. Using text analysis methods, we classify 360,000 articles in economics and 2.5 million in chemistry published between 1950 and 2020 into 90 and 80

topics respectively, each characterised by a coherent set of themes and vocabulary. We document three patterns common to both disciplines. First, topics tend to follow a rise-and-fall life cycle: their publication share increases, peaks, and then declines. Second, aggregate scientific activity keeps growing, both overall and within each topic, without any apparent loss of quality. Third, this growth comes with a constant "reset" in the topic composition of each discipline, as new combinations of topics keep emerging. Altogether, our findings bear on the debate over whether ideas are becoming harder to find.

Keywords

Economics of science; Public economics; Bibliometrics; Academic freedom; Hungary; University autonomy; Research productivity; Text analysis

SOMMAIRE

REMERCIEMENTS . . . . . **iii**

RÉSUMÉ ET MOTS CLÉS . . . . . **v**

ABSTRACT AND KEYWORDS . . . . . **vii**

SOMMAIRE. . . . . . **ix**

TABLE DES TABLEAUX . . . . . **xi**

TABLE DES ILLUSTRATIONS . . . . . **xiii**

GENERAL INTRODUCTION **1**

CHAPITRE 1 — HOW TO SILENCE RESEARCHERS? EVIDENCE FROM ILLIBERAL POLICIES IN HUNGARY **9**

I — Introduction 9
II — Political Context, Illiberalism and Threats to Academic Freedom 13
III — Academic Data Collection 15
IV — Empirical Strategy. 18
V — How Do Political Pressures Affect Scientists' Productivity? 22
VI — Academic Strategies under Political Pressure. 30
VII — What Is the Global Effect of Illiberal Policies? 39
VIII — Conclusion 46

CHAPITRE 2 — CHARACTERISING THE GAINS TO AUTONOMY IN THE PUBLIC SECTOR: EVIDENCE FROM FRENCH ACADEMIA **47**

I — Introduction 47
II — Context and Content of the Reform. 51
III — Data and Methodology. 53
IV — How Did the Reform Affect Institutions? 56
V — How Did the Reform Affect Researcher Productivity? 62
VI — Discussion 65
VII — Conclusion 68

CHAPITRE 3 — THE LIFE CYCLE OF IDEAS **69**

I — Introduction 69
II — Data and Method. 71
III — Implementation of the Algorithm 76
IV — Stylized Facts on the Lifecycle of Ideas in Economics and Chemistry 77
V — Conclusion 84

BIBLIOGRAPHIE **85**

Ouvrages imprimés 85
Chapitre dans un ouvrage imprimé. 86
Rapports imprimés. 86
Travaux universitaires 88
Articles de périodiques imprimés 89
Communication dans un congrès 96
Autres sources 97

ANNEXE 1 — ANNEXES DU CHAPITRE 1 : HOW TO SILENCE RESEARCHERS? EVIDENCE FROM ILLIBERAL POLICIES IN HUNGARY 98

1.A — Political Background and the State of Academia in Hungary 98
1.B — Data Collection and Harmonization Process 104
1.C — Academics' Political Alignment 109
1.D — Theoretical Framework: Exit, Voice, Loyalty or Self-censorship 121
1.E — Self-censorship and Research Direction. 129
1.F — Political Alignment and Impact on Productivity: Complementary Results 133
1.G — *Exit* and *Voice*: Complementary Results 147
1.H — Global effect: Complementary Results 154
1.I — Robustness Exercises 156

ANNEXE 2 — ANNEXES DU CHAPITRE 2 : CHARACTERISING THE GAINS TO AUTONOMY IN THE PUBLIC SECTOR: EVIDENCE FROM FRENCH ACADEMIA 165

2.A — Appendix to Section II 165
2.B — Appendix to Section III 166
2.C — Appendix to Section IV 170
2.D — Appendix to Section V 179
2.E — Appendix to Section VI 180

ANNEXE 3 — ANNEXES DU CHAPITRE 3 : THE LIFE CYCLE OF IDEAS 186

3.A — Topic Evolution Through Years. 186
3.B — Topic Weights and Classification. 189
3.C — Topic Popularity Trajectories 194
3.D — Data Collection and Treatment 197
3.E — Details on Topic Identification 200
3.F — Optimisation of the Embedded Topic Model 201
3.G — Alternative measures 204

TABLE DES MATIÈRES. 206

TABLE DES TABLEAUX

Table 1.1 : Effect on Academic Productivity - Targeted Topics 26
Table 1.2 : Trends in Outcomes for Political Opponents, by Public Affiliation 28
Table 1.3 : Allies and Opponents: Effect on Network and Influence 29
Table 1.4 : Collaboration and Influence - Opponents Network 30
Table 1.5 : Political Opponents Network: Migration and International Collaboration . . . 34
Table 1.6 : Political Opponents Network: Voicing Behaviours 35
Table 1.7 : Researchers Working on Gender: Research Similarity 38
Table 1.8 : Effect on Research Quantity, Quality and Network 41
Table 1.9 : Effect on International Collaboration, Migration and Exit 44
Table 1.10 : Aggregate Treatment Effects by Mandate 45

Table 2.1 : Effect of the Reform on Number of Researchers 58
Table 2.2 : Effect of the Reform on Productivity of Newly Affiliated Researchers 59
Table 2.3 : Effect of the Reform on Productivity at the Institution Level 60
Table 2.4 : Effect of the Reform on Productivity at the Individual Level 64
Table 2.5 : Effect of the Reform on Productivity at the Individual Level by Quantile of Citations to 2000-2002 Papers 66

Table 1.B.1 : Statistics on the Merging Procedure between OpenAlex and National Repositories 108
Table 1.C.1 : Pre-Treatment Descriptive Statistics by Group. 113
Table 1.E.1 : Pairwise Similarity of Publications Across Groups, 2000–2009 130
Table 1.F.1 : Effect on Academic Productivity - Allies 133
Table 1.F.2 : Effect on Academic Productivity - Political Opponents 133
Table 1.F.3 : Effect on Research Quantity and Network – Political Opponents 134
Table 1.F.4 : Effect on Research Quality – Political Opponents. 134
Table 1.F.5 : Effect on Academic Productivity - Network of Political Opponents 135
Table 1.F.6 : Effect on Academic Productivity - Network of Allies 135
Table 1.F.7 : Effect on Academic Productivity - Climate 136
Table 1.F.8 : Trends in Outcomes for Political Opponents, Gender-Related Topics and Political Opponents in Gender Topics 139
Table 1.F.9 : Trends in Outcomes for Political Opponents, Sensitive Topics and Political Opponents in Sensitive Topics 140
Table 1.F.10 : Trends in Outcomes for Political Opponents, by closeness to pre-2010 ERC recipients. 141
Table 1.F.11 : Trends in Outcomes for Political Opponents, by closeness to any-period ERC recipients 141
Table 1.F.12 : Placebo: Public Affiliation among Researchers Unrelated to Opponents or Allies 142
Table 1.F.13 : International Collaborations - Opponents 143
Table 1.F.14 : International Collaborations - Allies 144
Table 1.F.15 : Network of Allies: Effect on Network and Influence 144
Table 1.G.1 : Effect on *Voicing* behaviours – Political Opponents Treatment 147

Table 1.G.2 : Ally Network: Migration and International Collaboration 148
Table 1.G.3 : Ally Network: Additional Measures 149
Table 1.G.4 : Political Opponent Network: Additional Measures 150
Table 1.G.5 : Triple Difference-in-Differences: Researchers working in Gender 152
Table 1.G.6 : Triple Difference-in-Differences: Researchers working on Sensitive Topics 152
Table 1.G.7 : Researchers Working on Sensitive Topics: Research Similarity 153
Table 1.H.1 : Effect on Migration, Hungary vs. Central Europe 154
Table 1.H.2 : Effect on International Migration - By Type of Affiliation 154
Table 1.H.3 : Effect on International Migration - By Region of destination in Europe . . . 155

Table 2.B.1 : Descriptive Statistics - Institution-level Outcomes pre-2007. 166
Table 2.B.2 : Descriptive Statistics - Institutions by Domain 167
Table 2.B.3 : Descriptive Statistics - Institution-level Characteristics 168
Table 2.B.4 : Descriptive Statistics - Researcher-Level Outcomes pre-2007. 168
Table 2.B.5 : Descriptive Statistics - Researchers by Domain 169
Table 2.C.1 : Effect of the Reform on Number of Researchers 170
Table 2.C.2 : Effect of the Reform on Number of Researchers - Data Truncated up to 2019 172
Table 2.C.3 : Effect of the Reform on Number of Stayers and Movers 172
Table 2.C.4 : Effect of the Reform on Number of Researchers by Domain of Institution . 177
Table 2.C.5 : Effect of the Reform on Productivity of Affiliated Researchers by Domain of Institution 177
Table 2.D.1 : Effect of the Reform on Productivity at the Individual Level - Limiting the Period to 2019 179
Table 2.D.2 : Effect of the Reform on Productivity at the Individual Level Conditional on not Having Left Academia 180
Table 2.D.3 : Effect of the Reform on Productivity at the Individual Level Conditional on not Having Left Academia in 2025 180
Table 2.E.1 : Effect of the Reform on Number of Researchers - No Control Variables . . . 181
Table 2.E.2 : Effect of the Reform on Productivity at the Institution-Level - No Control Variables 182
Table 2.E.3 : Effect of the Reform on Productivity at the Individual Level - No Control Variables 184
Table 2.E.4 : Effect of the Reform on Number of Researchers - Never Treated Only 184
Table 2.E.5 : Effect of the Reform on Productivity at the Institution Level - Never Treated Only 184
Table 2.E.6 : Effect of the Reform on Productivity at the Individual Level - Never Treated Only 185

Table 3.B.1 : Distribution of Topic Weights by Rank and Topic Type. 190
Table 3.B.2 : Topic Labels and Keywords: economics 192
Table 3.B.3 : Topic Labels and Keywords: chemistry 193
Table 3.C.1 : Parameter Settings Across the Three Classification Rounds 195

TABLE DES ILLUSTRATIONS

Figure 1 : Number of Publications in Scopus by Country of Affiliation 2

Figure 1.1 : Opponent Network Example 21
Figure 1.2 : Effect on Publications, by Political Alignment. 24
Figure 1.3 : Effect on Hungarian Academics' Productivity, by Political Alignment. 24
Figure 1.4 : Effect on the Probability of Leaving Hungary 33
Figure 1.5 : Effect on the Probability of Leaving Hungary, by Field and by Returns to Migration 33
Figure 1.6 : Effect on the Number of *Rule of Law* Publications, in English 36
Figure 1.7 : Effect on Research Similarity with pre-2005 Publications - Gender and Sensitive Topics 37
Figure 1.8 : Effect on Publications in top 10% cited journals by field. 42
Figure 1.9 : Effect on Publications by Journal Rank and on the National Co-Author Network 43

Figure 2.1 : Timeline of the Reform up to 2020 53
Figure 2.2 : Effect of Administrative Autonomy on Number of Researchers 56
Figure 2.3 : Effect of Administrative Autonomy on Number of Researchers who First Published Abroad 57
Figure 2.4 : Effect of Administrative Autonomy on Total Production of Researchers with Less than 5 Years of Seniority at the Institution. 58
Figure 2.5 : Effect of Administrative Autonomy on Total Production of Researchers at Treated Institutions 59
Figure 2.6 : Effect of Administrative Autonomy on Number of Researchers by IDEX Wave of the Institution 61
Figure 2.7 : Effect of Administrative Autonomy on Publications, Citations and Publications in the Top 5% Cited Journals - Individual Level 63
Figure 2.8 : Effect of Administrative Autonomy on Publications and Novelty by Quantile of Citations to 2000-2002 Papers 65

Figure 3.1 : Topic Allocation Strategies 75
Figure 3.2 : Perplexity by Number of Clusters. 76
Figure 3.3 : Topic Evolution over Time 77
Figure 3.4 : Average Topic Trends, Relative to year of emergence 78
Figure 3.5 : Average Topic Trend According to Rank 79
Figure 3.6 : Annual Weighted Number of Publications per Topic 79
Figure 3.7 : Average Topic Trend Relative to Emergence. 80
Figure 3.8 : Average Topic Trend 80
Figure 3.9 : Average Topic Trend in Citations 81
Figure 3.10 : Average Topic Trend in Originality and Generality 81
Figure 3.11 : Average Topic Trend in Similarity to Backward and Forward References . . 82
Figure 3.12 : Number of New Long-Lasting Combinations 82
Figure 3.13 : Average Topic Trend: Originality and Similarity 83

Figure 1.A.1 : Academic Freedom Index – Central-Eastern European Countries (2000-2024) . . . 99
Figure 1.A.2 : Decomposition of the Academic Freedom Index - Hungary (2000-2024) . 100
Figure 1.A.3 : Total Number of PhD Students in Hungary between 2008 and 2019 – by Nationality . . . 100
Figure 1.A.4 : Total Number of Entering Authors between 1990 and 2022 in Hungary . . 101
Figure 1.A.5 : Total spent by the government on research by type of socio-economic objective . . . 102
Figure 1.A.6 : Number of Professors at the Central European University (1990-2025). . . 103
Figure 1.A.7 : Exit flows of professors from the Central European University (2000-2025) 103
Figure 1.B.1 : Total Publications per Year in National Databases . . . 109
Figure 1.B.2 : Average Publications per Year - Full Sample . . . 110
Figure 1.B.3 : Average Publication Trajectories per Author . . . 111
Figure 1.C.1 : Distribution of Network Index - Control Groups . . . 114
Figure 1.C.2 : Distribution of Network Index - Quantiles . . . 114
Figure 1.C.3 : Network Proximity – Effect on Influence and Collaborations . . . 115
Figure 1.C.4 : Distribution by domain of researchers working in gender topics in entire dataset, control group and treated group of the matched sample.. . . . 117
Figure 1.C.5 : Distribution by domain of researchers working in sensitive topics in entire dataset, control group and treated group of the matched sample.. . . . 118
Figure 1.C.6 : Distribution by domain of researchers working in climate-related topics in entire dataset, control group and treated group of the matched sample. . . . 119
Figure 1.C.7 : Number of entering authors working on gender topics. . . . 120
Figure 1.D.1 : Optimal Strategy in an "Exit, Voice, Loyalty or Self-censorship" Game . . 123
Figure 1.D.2 : Timing of Responses . . . 125
Figure 1.D.3 : Population Shares as a Function of Political Pressure . . . 127
Figure 1.E.1 : Journal Rank of Politically Sensitive and Other Publications . . . 131
Figure 1.E.2 : Share of Politically Sensitive Publications by Journal, Before and After 2010 132
Figure 1.F.1 : Number of ERC recipients per year per political status. . . . 138
Figure 1.F.2 : Number of Publications for ERC Recipients, Social Sciences, per Political Group . . . 138
Figure 1.F.3 : Ally Network and Opponent Network: Effect on Total Collaborations. . . . 145
Figure 1.F.4 : Ally Network and Opponent Network: Effect on New Collaborations 145
Figure 1.F.5 : Ally Network and Opponent Network: Effect on Citations . . . 146
Figure 1.G.1 : Effect on the Probability of Publishing in Gender - Triple DiD . . . 151
Figure 1.I.1 : Placebo Event Study, Within Hungary . . . 157
Figure 1.I.2 : Placebo Difference-in-Differences by Fake Treatment Year, Within Hungary 158
Figure 1.I.3 : Random Treatment Assignment, Within Hungary . . . 159
Figure 1.I.4 : Randomization Inference, Political Opponents . . . 160
Figure 1.I.5 : Placebo Exercises, Cross-Country . . . 162
Figure 1.I.6 : Random Treatment Assignment, Cross-Country . . . 163
Figure 1.I.7 : Randomization Inference, Cross-Country . . . 164

Figure 2.B.1 : Average Number of Authors in 2004-2007 by Domain of Research and Institution by Cohort of Treatment . . . 167
Figure 2.B.2 : Average Number of Researchers per Year in Treated and Control Institutions 168
Figure 2.B.3 : Average Scientific Production per Year in Treated and Control Institutions 168
Figure 2.B.4 : Log of Number of Researchers (left), Number of Institutions (right) and Average Citations (bottom) by French Department . . . 169
Figure 2.C.1 : Effect of Administrative Autonomy on Number of Researchers with Previous Affiliation Outside of France . . . 170
Figure 2.C.2 : Effect of Administrative Autonomy on Number of Researchers by Seniority 171
Figure 2.C.3 : Effect of Administrative Autonomy on Number of Researchers Retiring Each Year . . . 172

Figure 2.C.4 : Effect of Administrative Autonomy on Total Production of Researchers who First Published Abroad . . . 173
Figure 2.C.5 : Effect of Administrative Autonomy on Number of Researchers by Previous Affiliation . . . 173
Figure 2.C.6 : Effect of Administrative Autonomy on Citations of Researchers at Treated Institutions . . . 174
Figure 2.C.7 : Effect of Administrative Autonomy on Publications of Treated Researchers in Top 5% Cited Journals . . . 174
Figure 2.C.8 : Effect of Administrative Autonomy on New Phrase Combinations Weighted by Later Reuse in OpenAlex . . . 175
Figure 2.C.9 : Effect of Administrative Autonomy on Publications and Citations by IDEX wave of the Institution . . . 175
Figure 2.C.10 : Effect of Administrative Autonomy on New Phrase Combinations Weighted by their Reuse by IDEX Wave of the Institution . . . 176
Figure 2.C.11 : Effect of Administrative Autonomy on Total Publications in First-Wave IDEX Universities, by 2000-2002 Productivity . . . 176
Figure 2.C.12 : Effect of Administrative Autonomy on Number of Researchers in First-Wave IDEX Universities, by 2000-2002 Productivity . . . 178
Figure 2.D.1 : Effect of Administrative Autonomy on Novelty of Research . . . 179
Figure 2.D.2 : Effect of Administrative Autonomy on Publications and Novelty by IDEX Wave of the Institution . . . 181
Figure 2.E.1 : Effect of Administrative Autonomy on Total Researchers and Publications - No Control Variables . . . 182
Figure 2.E.2 : ATT by Cohort of Adoption . . . 183
Figure 2.E.3 : Effect on Researchers Based on Year of Treatment of Affiliation, Varying the Year in which Treatment is Assigned . . . 183

Figure 3.A.1 : Share of Publications and Trend in Economics, All Topics, Normalised Shares . . . 187
Figure 3.A.2 : Share of Publications and Trend in Chemistry, All Topics, Normalised Shares . . . 188
Figure 3.A.3 : Publication Share Trajectory Relative to Maximum, by Peak Decade . . . 189
Figure 3.A.4 : Publication Share Trajectory Relative to Emergence, by Peak Decade . . . 189
Figure 3.C.1 : Aggregate Trend, All Topics . . . 196
Figure 3.D.1 : Number of Total Publications in Sample, Per Domain . . . 197
Figure 3.D.2 : Average Number of Publications per Year per Researcher, by Domain . . . 198
Figure 3.D.3 : Total number of entries in sample and in entire database . . . 198
Figure 3.D.4 : Total Publications Count, by Sample . . . 199
Figure 3.F.1 : Bell-shaped topics in observed and simulated data . . . 202
Figure 3.F.2 : Coherence to perplexity ratio for all specifications in the three-subjects simulated data . . . 203
Figure 3.F.3 : Number of abstracts per simulated subject and simulated date (left) and by algorithm subjects and simulated date (right) . . . 204
Figure 3.G.1 : Average Semantic Distance to all OpenAlex Papers . . . 205

General Introduction

Scientific output has grown exponentially in the past decades (Bornmann and Mutz, 2015). The number of recorded publications has skyrocketed in all fields, and so has the number of authors. This growth has come with the emergence of concerns about the future of science: have all impactful ideas already been found (Gordon, 2000; Gordon, 2017)? Has the background knowledge scientists need to advance science grown into an overwhelming "burden" that limits their ability to push the frontier (Jones, 2009)? Science is also, to a large extent, publicly funded. As fiscal constraints tighten across Europe, the cost of funding scientific research becomes more salient. Are the returns to science worth a cost borne mainly by the public sector? Is it possible to produce more high-quality scientific papers without increasing public spending? In particular, is the recent shift toward competitive, grant-based funding in Europe efficient (Azoulay, Graff Zivin, and Manso, 2011; Azoulay and Li, 2020; Veugelers, Wang, and Stephan, 2025)? Finally, scientific institutions also bear the brunt of anti-intellectualism in a context of rising populism (Roberts Lyer, Saliba, and Spannagel, 2023). What does this entail for the production of ideas? What are the economic effects of policies implemented by populist parties on science? In this thesis, my co-authors and I attempt to address some of these new challenges for science.

In Aghion and Howitt (1992), the engine through which scientific ideas contribute to better living conditions is made endogenous through creative destruction: firms invest in research and development, and entrants make existing products obsolete, and their productivity gains benefit to the broader economy. Industrial R&D, which depends critically on the advancement of scientific knowledge, is thus at the core of the engine of growth, allowing the prospect of monopoly rents to be greater than the costs necessary for this enterprise. By contrast, the scientific activities on which R&D relies do not appear as explicitly in the framework. Scientists compete against each other, but they only capture a fraction of the monetary value of their discoveries. The gap between scientific and economic value is particularly visible in Europe, in which the science diffuses less to industry compared to the US (Draghi, 2024). Yet, Figure 1 shows that even as China's publication output skyrocketed to make it one of the largest scientific superpowers, Europe has maintained its place on the podium. Unlike innovators from Aghion–Howitt–type growth models who respond to monopoly rents, scientists are not only motivated by monetary payoffs, but also by prestige, freedom in their work, and their intrinsic interest in their research (Merton, 1957; Stern, 2004; Aghion, Dewatripont, and Stein, 2008; Roach and Sauermann, 2010; Balsmeier and Pellens, 2016; Arts and Veugelers, 2020). Measuring the output of scientific activities is also less straightforward than for R&D: the economic returns to scientific research, including when it is publicly funded, are large but hard to measure (Gazzani et al., 2026). Even projects that appear to produce little in the short term may pave the way for major discoveries with large economic impact (Bhattacharya and Packalen, 2020). But like growth, science does not arise from thin air. Publications are produced by scientists, usually in laboratories. These scientists are workers, who are affected in how they use their labour by the current state of their discipline, the organisation of the institutions they work in and the incentives they provide for scientific discoveries in their career, and the broader tendencies in the society they participate in, the attitudes towards science, and the funding that governments decide to invest in it (Stephan, 1996).

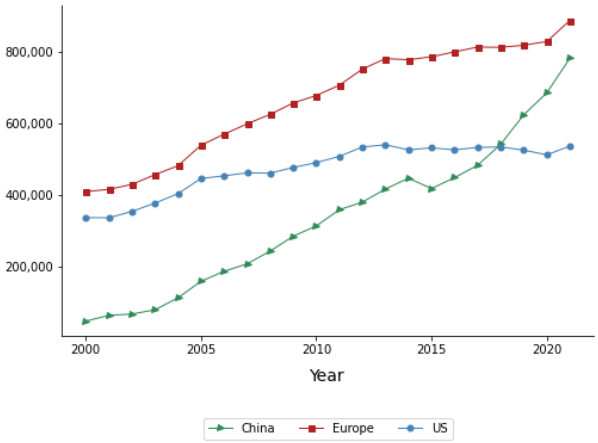


Figure 1 : Number of Publications in Scopus by Country of Affiliation

Notes: This figure represents the number of scientific publications each year in Scopus, by country of affiliation. Scopus is a bibliometric database released by Elsevier in 2004, which in 2024 covers 43,132 scientific journals, 78 million publications, and 16 million authors. The line labelled Europe includes the 27 countries of the European Union, the United Kingdom, Switzerland, Norway, and Iceland. Source: Aghion, Antonin, et al. (2026), joint work with Philippe Aghion, Céline Antonin, Luc Paluskiewicz, David Strömberg, Xueping Sun and Karolina Westin.

Each of the three chapters of this thesis contributes to a strand of the literature on the economics of science. The first chapter relates to the literature on the political determinants of science. The link between institutions and economic growth is well established (Acemoglu, Johnson, and Robinson, 2005). A recent study by Hager et al. (2026) assesses the importance of liberal democracy for the production of science in the long run, by comparing democracies and autocracies and analysing changes in scientific output during regime changes. The literature also analyses the effect of specific policies relating to academic freedom in autocracies and democracies. A large body of historical work studies the impact of forbidding scientists from doing science based on their characteristics or their beliefs, for instance when Nazi Germany dismissed Jewish scientists (Waldinger, 2012; Waldinger, 2016), or under the pressure of the Inquisition (Dewitte et al., 2022; Cox and Figueroa, 2025). The literature also studies the same question in more recent contexts: in 2025, threats by the US government to cut funding to projects containing a list of undesirable words have led to large changes in the direction of economists' research (Rohner, Vanden Eynde, and Widmer, 2026). The effect on the direction of activities of potential scientists of official or social pressure is also documented, both for historical and contemporary contexts: Blasutto and de la Croix (2023) estimate that censorship under the Catholic Church in early modern Italy lowered the average log publication per scholar by 43%, while Acemoglu, Yang, and Zhou (2021) show that in China, faculty members shift the direction of their research to match that of their new department leader, in particular when they are a member of the Communist Party. Barriers to the free flow of ideas and scientists also strongly reduce scientific output: Iaria, Schwarz, and Waldinger (2018) show that the collapse in international scientific exchanges after World War I reduced the ability of scientists to produce cutting-edge research, while Jia et al. (2024) and Aghion, Antonin, et al. (2026) show that the China Initiative, which limited cooperation between US and Chinese scientists, significantly decreased scientific output on both sides. Our contribution is threefold. First, we study not a dictatorship, but rather a more recent form of repression, illiberal democracy, and trace its consequences over four successive mandates rather than only its immediate effects. Second, we show that persecution of the few, even without the extreme punishments of the Inquisition, coincides with a deterioration of research output for all researchers in Hungary relative to neighbouring countries and increased scientific isolation, which we interpret as a chilling effect on the entire country. Finally, we study the reactions of affected scientists and document self-censorship more precisely than before, using text-based measures.

The second chapter contributes to the broad literature on the organisation of science and on how the structure of incentives affects the production of scientific output. This question relates to the large literature on the optimal provision of public services, regulation and incentive contracts, both theoretical (Baron and Myerson, 1982; Laffont and Tirole, 1986; Laffont and Tirole, 1987; Pint, 1991) and empirical (Wolak, 1994; Gagnepain and Ivaldi, 2002; Fabrizio, Rose, and Wolfram, 2007; Cooper et al., 2011; Gagnepain, Ivaldi, and Martimort, 2013; Gaynor, Moreno-Serra, and Propper, 2013; Bloom, Propper, et al., 2015). As funding science through grants has spread (Price, 2019; Azoulay and Li, 2020; Carnehl, Ottaviani, and Preusser, 2025), this literature has been combined with theoretical work on contests, the verification of costly signals and optimal participation (Moldovanu and Sela, 2001; Condorelli, 2009; Ben-Porath, Dekel, and Lipman, 2014; Adda and Ottaviani, 2024), even taking into account the cost of competing (Gross and Bergstrom, 2019). While individual-level competitive funding typically increases the scientific output of recipients (Azoulay, Graff Zivin, and Manso, 2011; Gush et al., 2018), it also tends to select against those with the most novel ideas and therefore incentivises all scientists to pursue less risky and more incremental research (Ayoubi, Pezzoni, and Visentin, 2021; Veugelers, Wang, and Stephan, 2025). Contests for funding at the level of clusters of universities only have moderate effects (Möller, Schmidt, and Hornbostel, 2016; Menter, Lehmann, and Klarl, 2018; Buenstorf and Koenig, 2020; Corsini, 2022; Corsini and Pezzoni, 2023); in the context of French universities, I find no difference between those that won the contest and those that did not. Here, rather than focusing on funding, I focus on the organisation of the workplace of scientists. Earlier work finds that more autonomous universities produce more and better research (Aghion, Dewatripont, Hoxby, et al., 2010), and that the systems that provide stability for French scientists encourage Matthew effects (Hall and Mairesse, 2025). I show that the relationship between administrative autonomy and science is not as positive in France: it increased output at the institution level by making universities attract more researchers, while lowering it at the individual level for researchers who were already hired.

The third chapter contributes to the effort to describe science as an institution that evolves according to predictable patterns (Dasgupta and David, 1994; Stephan, 1996), and that new paradigms periodically disrupt (Griliches, 1957; Kuhn, 1962; Popper, 1959). More specifically, it speaks to an active debate in economics: whether ideas have become harder to find in recent decades. Gordon (2017) argues that new technological revolutions are less impactful than the first revolutions of steam or electricity. Jones (2009) argues that the accumulation of knowledge raises the cost of reaching the frontier. Bloom, Jones, et al. (2020) show that the productivity of research and development per research worker has declined over the past decades in several domains. Within science, research teams have grown larger (Wuchty, Jones, and Uzzi, 2007), progress has slowed in fields that have become increasingly canonised as they expand (Chu and Evans, 2021), and both papers and patents have become less disruptive since the mid-twentieth century (Park, Leahey, and Funk, 2023). On the other hand, Mokyr (2017) argues that the origins of growth lie in institutions that are still active today, that major breakthroughs keep occurring (Mokyr, 2014) and that interactions between current scientific advances should be enough to sustain growth (Mokyr, 2018). Theory predicts that even if existing research lines are exhausted faster because of the large increase in the number of researchers, the entry of new paradigms, by leading incumbents to exit, "resets the innovation clock" and increases growth (Aghion, Bergeaud, et al., forthcoming). Drawing on this theoretical framework, we analyse descriptively the evolution of two scientific fields, economics and chemistry, to confront these theoretical findings with data and bridge the gap between the two sides of the debate. While economists have long studied publications in their own field, our approach differs in that we do not rely on JEL codes (Angrist and Pischke, 2010; Card and DellaVigna, 2013; Boppart and Staub, 2016) but on our own classification. Our paper is part of a new wave of the literature that uses methods developed in computer science to map economics, and explain how arguments are built in the discipline (Garg and Fetzer, 2026; Ash, Ciccarone, and Galletta, 2026) and to show how the flow of knowledge between topics influences the semantic content of papers (Harris, 2026). Our analysis is not only applied to economics, but also to chemistry, a physical science

that differs from economics not only in content but also in its dependence on funding (Bertolotti, Myers, and Tham, 2025).

This thesis attempts to contribute to our understanding of how scientific ideas are made, upstream of economic growth, and of how the conditions in which academics operate shape the development and the impact of these ideas. This effort is empirical: it relies on statistical analysis of the output of researchers in bibliometric data, in order to precisely measure the effect of these drivers of scientific activity. Using this data has many advantages. First, it allows me to describe precisely what researchers produce and to quantify the output of scientific endeavours: in which type of journal their work is published, how much impact it has, and what the choice of words reveals about the content and novelty of their research. Second, they let me retrace the choices that researchers made of where to locate, and with whom, using their affiliation histories. This in turn allows me to describe how their choices and output evolve when confronted with a shock, be it political, in Chapter 1, or organisational, in Chapter 2, or depending on the history of their discipline in Chapter 3. Finally, relying on open bibliometric data, in particular OpenAlex, supports the practice of open science, which is crucial to establish good incentives for the ethical practice of science and the diffusion of knowledge.

This thesis approaches science as sector of the economy. It is unique in its specificities, but it is also subject to the same forces that govern manufacturing or other services. The insights brought by the economics of science are therefore not limited to science itself, and all the chapters of this thesis speak to social issues that go beyond academic research. The effect of the decrease in political freedom in Hungary that we describe in Chapter 1 on the production and choices of academics, whether or not targeted by the regime, also has implications for the rest of the economy of illiberal countries. In particular, it causes more scientists to leave the country, affecting the stock of human capital much like more mundane policies such as taxes (Akcigit, Baslandze, and Stantcheva, 2016). It also distorts the direction towards which scientists put their effort, and similar distortions could not only be reflected in the choices of inventors of patents, but also of entrepreneurs and workers who may also face political pressures and uncertainty. Chapter 2 studies the effect of the French reform of 2007, which delegated more administrative autonomy to universities, on scientific output at the researcher and institution levels. Similar reforms have been implemented in other public services, in France and in other countries. While the analysis cannot speak directly to those other contexts, it can help to explain potentially similar effects in public services which, like academia, have a dual objective of producing efficiently and serving all citizens of the country equally. In Chapter 3, we use unsupervised topic modelling to analyse the evolution of two scientific fields, economics and chemistry. In doing so, we contribute to the broad debate about whether ideas have become harder to find as science progresses, a debate that extends beyond publications to research and development, and we explain how ideas renew themselves.

Chapter 1: How to Silence Researchers? Evidence from Illiberal Policies in Hungary

This chapter is co-authored with Luc Paluskiewicz.

The first chapter of the thesis focuses on the effect of the restriction of academic freedom on research output, strategies of researchers in relocation and subject of study, and the global standing of science in the country. As the Cold War was drawing to a close, Fukuyama (1989) predicted that with no clear ideological rival left for liberal democracies, threats to this global consensus would arise from democracy itself, and indeed, a new type of regime called "illiberal democracy" (Zakaria, 1997) has emerged since the 1990s. In this type of regime, leaders are typically still elected in "free but unfair" elections while restricting their citizens' freedom of expression. Scientists are typically among the prime targets of the government in illiberal democracies, and academic freedom is usually one of the first freedoms to be curtailed (Roberts

Lyer, Saliba, and Spannagel, 2023). We study the case of Hungary to assess the effects of this decline in academic freedom on the production of ideas and on choices of scientists in their research topics and affiliations. Hungary is a particularly relevant case to study illiberal policies: Viktor Orbán unexpectedly obtained a supermajority, receiving 53% of the vote and 68% of the seats when his party won the elections of 2010. He then used it to change the constitution of the country as well as the organisation of crucial areas of the state such as the judiciary and the media authority, in order to build a regime he himself described as an "illiberal state". He is also cited as a model by many far-right populists in Europe and North America, including Donald Trump, who in 2025 himself targeted academics soon after his return to power. To estimate the effect of political pressures on academics' output, but also on their individual responses and on the country as a whole, we build a dataset by combining and harmonising national bibliometric repositories from Hungary, Slovenia and Croatia, allowing us to cover the universe of publications and researchers in these countries, including publications in national languages. We also use data from OpenAlex, the open international bibliometric database that took over from Microsoft Academic Graph after it was discontinued in 2021. We then merge these databases, allowing us to follow researchers even after they leave their country.

We first address the question of political pressures by classifying researchers as perceived political opponents or allies based on their membership of, or proximity to, specific groups that were either designated or openly presented themselves as such, or based on their research topics prior to 2005. We then compare them to a subset of the "neutral" group of researchers who are not close to or members of such groups, matched on observable characteristics, in a difference-in-differences design around Orbán's return to power. Researchers who are perceived as opponents lose on almost every count, increasingly so with each new Orbán mandate: after 2010, their number of publications decreases by 18% each year compared to neutral researchers, in particular in the top 10% most cited journals in their field. This effect holds independently of language, accounting for publications in Hungarian and in English. This holds both for researchers who belong to particular targeted groups, such as members of the Central European University, and for those targeted due to the topics they work on, for instance those who study gender-related topics. We then investigate the channels through which this repression translates into worse outcomes. While we cannot entirely rule out that funding plays a role, our evidence suggests that it cannot explain the bulk of this effect: the penalty is similar in public and private institutions, and access to outside funding does not offset it. Instead, we find that pressure damages the network of affected researchers, as well as their influence. Those lost collaborations likely prevent them from producing more high-quality work. We also rule out the hypothesis that this trajectory comes only from being politicised, as we find no effect on perceived allies of the government.

We then turn to the responses of perceived opponents to these political pressures. To do so, we draw from the framework of "Exit, Voice, Loyalty" (Hirschman, 1972), which we extend in a static model to a situation in which researchers with varying returns to emigration and levels of dissent can choose to exit, voice, stay loyal or self-censor. Perceived opponents become about 5 percentage points more likely to hold an affiliation outside of the country for a period of at least four consecutive years, an exit concentrated among the researchers with the best pre-2010 outcomes, but they are not more likely to leave academia. They are not only passively enduring these pressures: our analysis of the content of the titles of their publications shows that they also "voice", expressing their dissent by publishing on democratic backsliding, the rule of law crisis or academic freedom. They however mainly do so in English, and they become less likely to voice in Hungarian over the successive mandates. Researchers who publish in targeted topics prior to 2010 also publish research that is less related to their 2000–2005 research than neutral researchers do after Orbán is elected. We interpret these results as evidence of self-censorship: researchers push back, but they are also partly silenced.

Finally, we focus on the researchers who are not perceived as allies or opponents and compare their trajectory with that of comparable scientists from Slovenia and Croatia matched on observable characteristics, in the same difference-in-differences framework as above. Surprisingly, neutral Hungarian researchers publish more than their Slovenian and Croatian counterparts after 2010; but this increase is only driven by publications in their national language, to the detriment of publications in well-cited international journals, which decrease by about 60% during the third mandate. Likewise, their number of co-authors within Hungary rises while their number of collaborations with researchers abroad drops compared to the control. Finally, even though only a subset of scientists are under active political pressure, neutral researchers become more likely to leave: they are 1.3 percentage points more likely than comparable control authors to hold a foreign affiliation for a period of four consecutive years. We conclude that illiberal policies have strong implications not only for the direct victims of political pressures, but also for the country as a whole: we estimate that they caused 1,700 additional exits from academia, about 11% of the treated researchers.

Chapter 2: Characterising the Gains to Autonomy in the Public Sector: Evidence from French Academia

The second chapter turns to the organisational conditions of the production of ideas. I study the 2007 French university reform ("*Loi relative aux Libertés et Responsabilités des Universités*", LRU), which transferred to universities some of the administrative competences that belonged to the national Ministry of Higher Education and Research, in order to foster excellence through competition between institutions. Growing international competition and rising costs of the system had rendered its inefficiencies salient. The government's answer was to decentralise the management of staff and assets, previously fully handled by the State, and to give universities discretion in the use of some of their resources, in particular allowing more leeway in hiring to attract better researchers from France and from abroad.

In France, universities are not the only producers of academic research. In the same fields, publishing in the same journals and often working in the same lab, researchers employed by institutions with a different administrative status were, by construction, unaffected by the reform. I compare universities and their researchers to their counterparts in these other institutions: the National Centre for Scientific Research (CNRS), nonprofits such as the Institut Pasteur, or *Grandes Écoles* such as École Polytechnique. Since the reform was adopted in a staggered manner, I use the estimator of Callaway and Sant'Anna (2021), which compares cohorts of institutions and researchers who were treated to those that are not yet treated or never treated. I build two samples from OpenAlex data for the period 2003–2025. The first is aggregated at the level of an institution by large scientific domain (life, social, physical and health sciences) and year, which covers all researchers publishing in that institution in a given year, and allows me to count them by type, such as tenure and gender, and previous affiliation. It also measures scientific output at the aggregated level, both total and controlling for the number of active researchers. The second sample is at the researcher-year level, restricted to researchers who were already active prior to the reform.

I first estimate the effect of the LRU at the institution level using the first sample. The reform appears to have met its intended goal of making French universities more attractive to researchers, both from France and from abroad. Compared to the control group, treated institutions increase their stock of active researchers by about 15% over the following 10 years, including researchers who first published in a country other than France, with about the same proportion. These newcomers are active researchers who do not only contribute to teaching or to new administrative tasks, in particular those who were affiliated outside France for their first publication: the number of publications by researchers who arrived in the institution less than

five years earlier rises by 15%, and by 40% for those who first published abroad. Overall, total research output, measured by publications, rises by 23% in the treated institutions after treatment. Publications per researcher, however, do not change, and neither does research quality. The effect on output therefore appears to be driven only by the increased size of treated institutions. Whether the institution had a strong *Initiative d'Excellence* (IDEX) project, rewarded by early access to this competitive funding, does not appear to matter: those that received an IDEX early, a sign that the international jury assessing their project found it convincing, those that received it later and those that never received it follow similar trends. Scientific domains are however affected differently, life sciences benefitting the most from the reform.

I then turn to the effect of the reform on the individual researchers who were already active in treated institutions. Another way to raise the measured performance of universities is to make every researcher more efficient in their production, a mechanism for which the LRU provided tools such as excellence premia or allowing the possibility of raising wages through university foundations. I assess whether the reform increased individual productivity by estimating the same specification as above on the second sample, in which a researcher's treatment cohort is that of the institution they were affiliated with at the time the law passed. I find that on average, researchers in treated institutions have 1.5 fewer publications per year compared to the control, about 40% of their pre-reform average. Quality measures also decline during the first 8–10 years of treatment, before rising again; but the positive part of the effect does not survive using only data up to 2020, after which OpenAlex's data collection practices change. Even with data up to 2025, the negative part of the curve dominates the average effect over the whole period for most quality measures. About 31% of the effect on publications appears to be explained by exit from research; the majority of the effect thus comes from researchers who keep publishing every year. Yet again, the timing of the IDEX, which should be correlated with the quality of the excellence project, does not matter. Instead, the main driver appears to be research intensity, effects being more negative for researchers in the lowest quartile of citations received by papers published in 2000–2002. I interpret these effects as resulting from an increase in additional non-research tasks that divert researchers' effort away from research, as predicted by the multitasking framework of Holmstrom and Milgrom (1991). Teaching, university-level administration and the increased reliance of research on competitive funding which requires more work to secure are all plausible candidates, but I leave this disentangling for future work.

Chapter 3: The Life Cycle of Ideas

This chapter is co-authored with Philippe Aghion, Antonin Bergeaud, Luc Paluskiewicz and Gaëtan de Rassenfosse.

The third chapter relates to the debate about whether ideas have become harder to find over time. Gordon (2000) argues that the more recent technological revolutions have had less impact compared to the great inventions of the nineteenth century. Recent empirical evidence also shows that returns to research and productivity have declined in several sectors of the economy (Bloom, Jones, et al., 2020), and that research has become less disruptive in the past decades than in the mid-twentieth century (Park, Leahey, and Funk, 2023). Conversely, Mokyr (2014) argues that technological breakthroughs will keep occurring at the same pace. Recent theoretical work shows that even if ideas do become harder to find within existing research lines, the arrival of new lines should sustain growth (Aghion, Bergeaud, et al., forthcoming). We contribute to this debate with new empirical evidence that ideas only appear to become harder to find in sciences within topics but that new topics keep emerging, using topic modelling to map the evolution of economics and chemistry, respectively the least and most funding-dependent fields per the estimation of Bertolotti, Myers, and Tham (2025), over the period 1950–2020.

We analyse the content of scientific papers by applying the Embedded Topic Model (Dieng, Ruiz, and Blei, 2020) on the abstracts of over 360,000 journal articles in economics and 2.5 million journal articles in chemistry published between 1950 and 2020. The method is text-based, unsupervised, and allocates each paper to all topics based on their relevance to the paper; we establish the number of topics using perplexity on held-out data. We validate the results in economics through our expert knowledge. Overall, we classify all these abstracts into 90 topics for economics and 80 in chemistry, and describe their evolution, each one corresponding to a research subject or a method.

Using articles from economics and chemistry, this chapter documents three results that hold in both disciplines. First, a topic's importance, measured by its share of new articles each year, follows a bell-shaped path: it rises as the topic grows faster than its field, then declines as it falls behind the field-wide trend, more rapidly in economics than in chemistry. This is consistent with Gordon (2000)'s "fruit tree" parable: the easiest ideas within a topic are harvested first, and the remaining ones become harder to find. Second, total publication output grows steadily over the period, and we find no evidence that quality declines once a topic has peaked, so aggregate production shows no sign of diminishing returns. Third, new topic combinations emerge continually, which amounts to a perpetual "reset" in the production of ideas. This fits the absence of a slowdown in aggregate publication growth and supports Mokyr (2014)'s more optimistic view regarding the future of research.

Chapitre 1
How to Silence Researchers? Evidence from Illiberal Policies in Hungary

Ce chapitre reprend l'article « How to Silence Researchers ? Evidence from Illiberal Policies in Hungary », coécrit avec Luc Paluskiewicz, dont c'est le job market paper. [1]

I — Introduction

Explanations of long-run growth generally point to a small list of fundamental causes: luck, geography, culture and institutions (Acemoglu, 2009). This paper studies how institutions, a society's "rules of the game" (North and Thomas, 1973; North, 1990), shape the production of scientific knowledge. Several features of the institutional environment matter for the production of new ideas: the distribution of political power, which shapes incentives to innovate (Acemoglu, Johnson, and Robinson, 2005); the free and low-cost circulation of ideas (Mokyr, 2002); and a culture that values useful knowledge (Mokyr, 2017). Because scientific knowledge and new ideas are the main drivers of long-run growth (Romer, 1990; Aghion and Howitt, 1992), political shifts that concentrate power and restrict individual freedoms, including those of scientists, may threaten not only the production of ideas but also long-run development.

The fall of the Soviet Union and the end of the Cold War fostered the common expectation that liberal democracy would spread and eventually prevail (Fukuyama, 1989). Yet many democracies have since seen the rise of what Zakaria (1997) called "illiberal democracies": democratically elected regimes that continue to organise elections while restricting civil liberties and weakening the rule of law. In 2014, Viktor Orbán, Prime Minister of Hungary, openly embraced this term, which thus became a self-description rather than only a pejorative label used by critics.[2] The Hungarian case matters beyond its borders. Several right-wing and far-right

[1] We wish to express our particular gratitude to Philippe Aghion, Sascha Becker, Zsolt Boda, Gergely Kováts and Fabian Waldinger. We are also grateful to Antonin Bergeaud, Simon Bunel, Maxime Gravoueille, Adam Szeidl, Gaétan de Rassenfosse, Alexandra Roulet, Monika Schnitzer, David Strömberg, Erina Ytsma and Natalia Zinovyeva, as well as all participants at the NBER Summer Institute, EEA Annual Meeting, ETH, CERGIC, BSE Summer Forum, EUI PhD Workshop 2026, EAYE 2026, the PEPE Warwick Seminar, Applied Young Economist Webinar, INSEAD and Collège de France seminar, LMU Innovation Seminar, for their helpful comments and suggestions. We also thank María José González-Fuentes, Lena Greska, Robert J. Johannes, Corentin Laffitte, Milla Tarpola for their comments and help. The authors would like to acknowledge the financial and institutional support provided by the Paris School of Economics, Collège de France and Farhi Innovation Lab.

[2] The term *illiberal* first appeared in the literature to describe developments in South-East Asia (Bell et al., 1995), before it came into wider use after the publication of Zakaria (1997). Orbán explicitly described his governance model as an "illiberal state" in July 2014, in a public speech at the 25th Bálványos Summer Free University and Student Camp (see Viktor Orbán's speech). In the Hungarian context, we expand on detailed qualitative descriptions of regime change (Bánkuti, Halmai, and Scheppele, 2012; Boda et al., 2015; Halmai, 2018; Buzogány, 2020) and more particularly of the constraints on academic freedom (Kováts, 2015; Kováts and Rónay, 2021; Kováts, Derényi, et al., 2024). See Section II for more details.

leaders, including in France, Italy and Germany, have explicitly presented Orbán's Hungary as a model (Camus, 2022), and other governments have introduced similar restrictions. For instance, in the United States, the second Trump administration has cut funding to universities and research projects on political grounds, drawing on Orbán's policies (Secen and Kılınç, 2026). While the political dimensions of this shift are well documented, we know less about the consequences of illiberal policies for science and innovation.

This paper examines the consequences of political threats and declining academic freedom for the production of ideas, with a particular focus on the mechanisms through which academia and scientists are affected. Hungary under Orbán (2010–2026) provides a unique quasi-natural experiment setting to measure these consequences in a contemporary democracy.

Our analysis progresses through three successive and complementary research questions. First, how does political pressure affect the productivity of researchers targeted by the government? Second, how do researchers respond when targeted by political threats? Finally, what is the overall impact of illiberal policies on scientific production, beyond the researchers and fields that are direct political targets?

To answer these questions, we use rich bibliometric data that allow us to measure scientific outcomes for researchers based in Central European countries. Our primary dataset is constructed from the national research repositories of three Central European countries, namely Hungary, Slovenia and Croatia. These repositories, known as Current Research Information Systems (CRIS), are national bibliographic databases in which researchers and their institutions record their publications, in all languages and document types. They provide near-complete coverage of all research activities of researchers working in these countries. We translate and harmonise them to create a comprehensive dataset of scientific publications and researchers in these countries. In addition, we rely on OpenAlex, an international, centralised database with detailed information on co-authorship networks and institutional affiliations, which allows us to analyse outcomes that cannot be observed in the national data, in particular international collaborations and researcher mobility. From these two types of sources, we construct a broad author-year-level panel dataset covering the period from 1990 to today, allowing us to track academic outcomes over several decades.

We address the first question by studying whether, within Hungary, researchers are affected differently depending on their perceived political positions. Our empirical strategy relies on our ability to identify the perceived political alignment of researchers using information prior to Viktor Orbán's election in 2010, in particular their membership of, or proximity to, groups and institutions with a known political orientation, and the topics they work on. We then split academia into three groups. *Ally* researchers are perceived as supportive of the government, because they belonged before 2010 to political groups or institutions close to it. *Perceived political opponents* include two subgroups: *direct political opponents*, i.e., researchers who belonged before 2010 to political groups or institutions perceived as opposed to the government, and researchers who work on topics at odds with the government's agenda, such as gender studies. *Neutral* researchers, who represent the majority of Hungarian academia, have no collaboration ties with either aligned or perceived opponents. These categories reflect how researchers are likely to be perceived, not their actual political opinions or engagement. To account for spillovers, we also build a continuous measure of proximity to political opponents in the collaboration network, which allows us to study researchers who are not directly targeted but collaborate with them. Then, using a difference-in-differences design, we compare perceived opponents and allies with neutral ones. We find that researchers who are perceived as opponents experience a significant drop in their total number of publications, in their number of publications in top journals, and in the citations they receive. After 2010, political opponents publish 18% fewer papers than neutral researchers, and 59% fewer papers in the top 10% most-cited journals.

These findings are consistent across both subgroups of perceived opponents, political opponents and researchers working on sensitive topics.

We investigate the mechanisms through which this drop in academic productivity occurs. Funding does not appear to be the main channel: the penalty we measure on academic outcomes is similar for researchers affiliated to public and private institutions, and access to outside funding does not offset it. Instead, we show that the decline is likely due to the negative impact on researchers' networks: members of the network of political opponents have 16% fewer co-authors and form 15% fewer new collaborations after 2010. Conversely, researchers who are perceived as politically aligned do not experience the same decrease in their productivity. All these effects intensify over the years and across Viktor Orbán's successive terms, suggesting that the stronger the pressure on researchers becomes and the longer it lasts, the larger the consequences for their performance. These consequences are not confined to the core network of perceived opponents: their co-authors also publish 3% fewer papers and 12% fewer papers in top 10% journals, although the effect fades as the distance from clearly identified political opponents increases.

Second, we show that researchers perceived as political opponents have strong behavioural responses to political pressure. We draw explicitly on the *Exit, Voice, and Loyalty* framework of Hirschman (1972), adapted to the context of academic freedom backsliding, in which scholars facing political pressure may choose to adopt different strategies. We first develop a static theoretical framework, in which each researcher is characterised by two parameters, her return to migration and her political alignment, and we derive empirical predictions on researchers' choice of strategy. We then show that political opponents are more likely to leave Hungary and to obtain affiliations abroad, without leaving academia. This exit is concentrated among the researchers who are targeted by the government and who are at the top of the distribution in terms of pre-2010 academic outcomes. We also analyse the content of publications and show that these researchers are more likely to *voice*, that is, to publicly express dissent in their academic work, for instance by publishing on democratic backsliding, academic freedom or the rule of law. This *voicing* behaviour is strategic: researchers express their dissent mainly in English and reduce it in Hungarian. We also document *self-censorship* as a fourth strategy, distinct from both *voice* and *loyalty*. Scholars working on politically sensitive topics, such as gender studies, move away from their previous research topics in their Hungarian-language publications, while maintaining them in their English-language publications. They do so without converging towards the research of government-allied researchers. Like strategic *voicing*, self-censorship thus reduces the domestic visibility of politically sensitive research rather than abandoning it.

Finally, we assess whether declining academic freedom also has a cost for researchers who are identified neither as political opponents nor as allies. We compare neutral Hungarian researchers with matched researchers in two neighbouring Central European countries, Slovenia and Croatia, using a difference-in-differences strategy. This comparison captures the effect of the policy environment common to all Hungarian researchers, whereas the within-Hungary analysis captures the additional effect of being perceived as an opponent.Surprisingly, the main effect of illiberal policies on Hungarian academia as a whole is an increase in the number of publications, mostly in Hungarian. At the same time, Hungarian researchers become increasingly isolated: their national collaborations expand while their international collaborations contract. The quality of their research also declines, as their publications in the top 10% most-cited journals fall by about 60% during Orbán's third mandate. Finally, researchers working in Hungary before 2010 become more likely to move abroad durably and to leave academia, by about 25 to 30% relative to their pre-2010 baseline. These effects are not confined to the social sciences: we also find a decline in research quality and more exits among researchers in the life and health sciences. Our results suggest that illiberalism in Hungary did not reduce the number of publications, but changed how the same researchers do research: they publish more in Hungarian, less in top

international journals, and collaborate more at home than abroad, and that exit, either abroad or out of academia, is an increasing alternative for some of them.

Our work contributes to several strands of the literature.

First, building on a broad literature linking political institutions and economic growth (Acemoglu, 2009; Papaioannou and Siourounis, 2008; Funke, Schularick, and Trebesch, 2023), prior work has studied the impact of political shifts on science and innovation. Hager et al. (2026) show the importance of academic freedom and democratic institutions in shaping the creation of knowledge in the long run. From a historical perspective, political conflicts or pressure have also had a large impact on researchers' academic performance, as during the First World War (Iaria, Schwarz, and Waldinger, 2018), the Second World War and the Nazi regime (Waldinger, 2012; Moser, Voena, and Waldinger, 2014; Waldinger, 2016), and the Inquisition (Dewitte et al., 2022; Cox and Figueroa, 2025). Prior work has also measured the impact of Catholic censorship on scientific publication, for instance in early modern Italy (Blasutto and de la Croix, 2023) or on the mobility of "thinkers" (Becker, Pino, and Vidal-Robert, 2021). China has attracted particular attention, as it seems to reconcile tight political control over science with fast-growing innovation (Bergeaud and Verluise, 2022), although its academia has been affected by US sanctions (Aghion, Antonin, et al., 2026). Acemoglu, Yang, and Zhou (2021) show that, through changes in university leadership, Communist Party leaders exert pressure on Chinese universities, significantly changing research direction, while censorship has had lasting effects (Bai, Jia, and Yang, 2024). Beraja et al. (2023) highlight the reverse relationship, showing how innovation can in turn reinforce autocratic power. More recently, and in a democratic context, Rohner, Vanden Eynde, and Widmer (2026) find that economists respond strongly to political threats to funding after the US government published a list of undesirable words in 2025. Our paper contributes to this literature on the influence of politics on science and innovation in three ways. First, rather than an authoritarian dictatorship, we focus on a more contemporary form of repression, illiberal democracy, and explore its effects over a longer time horizon and across a wider set of domains. Second, we extend the analysis of the innovation costs of political repression from a small targeted group to the entire population of researchers within a country. Third, we provide an empirical application of the Hirschman (1972) framework in a contemporary political context: we precisely measure strong reactions from researchers and document new types of behaviour among scientists, such as expressing dissent and self-censoring.

Second, we contribute to the growing literature on the costs of political engagement for researchers. Aside from research on trust in science and political opinion (Cologna et al., 2025), our work is closely related to Alabrese, Capozza, and Garg (2024), who quantify the cost in credibility incurred by researchers who publicise their political opinions online. Here, we quantify the productivity costs for researchers whose political opinions are publicly known, or merely inferred from their affiliations or research topics: in our context, researchers do not need to express political views to bear these costs, as being perceived as politically opposed is enough.

Third, we build on the literature on the location choices and high-skilled migration of researchers and inventors to assess the effect of the regime change on migration behaviour. A large strand of this literature focuses on the impact of high-skilled immigration on the destination country or university (Borjas and Doran, 2012; Moser, Voena, and Waldinger, 2014; Lerner et al., 2024). Another strand has also examined the determinants of emigration from the country of origin, including women's rights (Nejad and Young, 2014), rising taxes (Akcigit, Baslandze, and Stantcheva, 2016), and network connections (Choudhury, Ganguli, and Gaulé, 2023; Becker, Lindenthal, et al., 2024). A related body of work has instead considered gains for both sides, in particular for the country of origin, through knowledge or productivity diffusion (Bahar et al., 2022; Batista et al., 2025; Prato, 2025). Our paper lies at the intersection of these last two strands:

for contemporary Central European countries, we quantify both the political determinants of emigration and its costs for scientific production in the country of origin.

Finally, our paper builds on the science of science literature that measures scientific outcomes. Prior work has studied the impact of external determinants on researchers' output (Azoulay, Graff Zivin, and Wang, 2010; Azoulay, Graff Zivin, and Manso, 2011; Aghion, Jackson, et al., 2023; Carayol, Henry, and Lanoë, 2025), both in terms of quantity and quality of their research. We also rely on previous scientometric work on academic outcomes (Kronegger, Ferligoj, and Doreian, 2011) and on work evaluating international databases (Boyack et al., 2014) and national research information systems, notably in Central Europe (Hladík and Renisio, 2025). We contribute to this literature in two ways. First, while Kwiek (2016) shows that top European researchers are comparable across countries, we document that political choices can have lasting effects that make researchers systematically different from those in countries where such choices were not made. Second, unlike previous comparisons of global and national systems (van Leeuwen, Van Wijk, and Wouters, 2016; Háló and Demeter, 2023), we are, to our knowledge, the first to combine an international database with several national ones across all disciplines, at the level of both researchers and publications. This allows us to cover non-English publications and the social sciences and humanities, which are usually excluded because international databases cover them poorly, yet are the primary targets of anti-intellectual rhetoric.

The remainder of the paper is organised as follows. Section II describes the political and academic context of Hungary after Viktor Orbán's election in 2010. In Section III, we detail our data collection work and define the variables we use in the analysis. Section IV details the empirical strategy we implement. Section V presents our first set of results, explaining to what extent and how political pressure affects scientific outcomes for researchers in Hungary. Section VI presents our second set of results, investigating the behavioural responses of scientists to political threats. Section VII presents our results regarding the global impact of illiberal policies on researchers. Finally, Section VIII concludes.

II — Political Context, Illiberalism and Threats to Academic Freedom

During the general election of April 2010, Viktor Orbán's party, the Fiatal Demokraták Szövetsége (*Alliance of Young Democrats*, Fidesz), unexpectedly obtained a supermajority, receiving 53% of the vote and 68% of the seats in Parliament, defeating the incumbent socialist majority. These electoral results marked a turning point in Hungary's modern political history. At the time of the election, Fidesz proposed a right-wing and conservative program aimed at creating a "new social contract for the country." The party's large electoral victory was interpreted by its leaders as a general rejection of liberal values by the Hungarian people, opening the way for structural changes in the state's organization to reflect new conservative values (Pap, 2017). Viktor Orbán, as leader of Fidesz, became Prime Minister of Hungary in May 2010. His government quickly began implementing reforms aligned with what he described as "non-liberal" principles.

According to several international organizations and NGOs, there are multiple dimensions through which post-2010 governments have facilitated Hungary's transition from a "liberal" to an "illiberal" democracy.[3] A new Fundamental Law is promulgated on April 25, 2011, which weakened several components of the system of checks and balances that previously limited governmental power. In several independent authorities (in particular the judiciary and the media authority), the Parliament dismissed existing leadership and replaced it with party affiliates, while expanding the authority's powers and making the removal procedures for their new

[3] Human Rights Watch or Amnesty International consistently report on threats to the rule of law in Hungary and the threats to democratic institutions.

heads significantly more restrictive (Bánkuti, Halmai, and Scheppele, 2012). More generally, since the 2014 general elections, Hungary's electoral process has been characterised as "free but unfair" (Bíró-Nagy, 2017), particularly due to restrictions on political advertising that have created unequal conditions for electoral competitors.[4] For instance, Kelemen (2017) highlights the chilling effect of Article 81 of Act CLXXXV on Media (2010), which grants regulators broad discretionary authority to sanction journalists whose content is deemed not "balanced, accurate, thorough, objective, and responsible". The European Commission for Democracy through Law has similarly criticised this law as overly broad and discretionary.[5]

Throughout all of the steps towards illiberalism, academic freedom in particular was increasingly targeted through different means. While its state had been constant since the fall of the Eastern Bloc, Hungary experienced a sharp decline in academic freedom after 2010.[6] In this paper, we consider that academic freedom has two dimensions. First, we use the researcher-centric definition given by Russell (1993), who defines it as the "freedom to question and test and develop new ideas without undue harassment or interference". In addition, we also consider a university-centric definition, from a historical perspective, in which we define academic freedom as the institutional right to autonomy for academic institutions. While we can consider the first dimension to have been continuously threatened since 2010, this section focuses on three salient reforms during which the second was breached, all linked to electoral cycles. First, while the new Act on Higher Education in 2011 was already a major turning point[7], it was replaced in 2014 by another method of control of universities through a new position, still appointed directly by the Prime Minister. Chancellors have the responsibility of managing the university's financial resources, which undermines the power of rectors and allows the State to maintain a certain level of control over the university's choices. Moreover, up to a third of the appointed chancellors have a connection to the ruling party (Kováts, 2015). Although the chancellor's power was somewhat checked by the introduction of a supervisory board to which both the rector and the chancellor belonged in 2015, their introduction led to an outcry in public universities, as it amply limits the power of rectors and muddies the line between university management and the government (Rónay, 2018). This process was furthered by the introduction of consistories in 2015, financial boards in which four out of five members were named by the government. Second, as part of the campaign leading up to the 2018 legislative elections, the Central European University (CEU), founded and funded by billionaire George Soros, was targeted as early as late 2017. This led to a legal battle that forced the university to permanently relocate to Vienna in 2019. Directors and more generally all professors affiliated with the University were accused of being both agents of external forces and of "social degeneracy", especially due to the existence of gender studies classes (Enyedi, 2022). The final and largest changes imposed on public higher education by the government has been the "regime change" started in 2018, when the Budapest University of Economic Sciences, now called Corvinus University Budapest, became owned by a public interest foundation. This model was then applied to twenty other public universities comprising the majority of university students in the country by 2022. While this seems to relax the State constraints on universities, the boards of these public interest universities comprise many Fidesz-aligned trustees (Szent-Iványi and Tétényi, 2022), meaning that up to a third of these trustees are Fidesz politicians, not including others who are supportive of the party (Kováts, Derényi, et al., 2024).

[4] See the OSCE / ODIHR Limited Election Observation Mission report.

[5] Opinion No. 798/2015 of the Venice Commission.

[6] This is particularly striking when considering the Academic Freedom Index (AFI) developed by V-Dem for Hungary. While it is constant between 1991 and 2010 at 0.95, it drops to 0.63 already in 2012 and keeps declining down to 0.3 in 2024, the lowest level among all the members of the European Union. See Figure 1.A.1 and Figure 1.A.2.

[7] Act CCIV of 2011, in which "autonomy" and "freedom" were no longer guaranteed but replaced by the necessity of "the spiritual and intellectual renewal of the nation". Furthermore, university rectors, who are responsible for professor recruitment, formerly elected by their peers, had to be appointed by the Prime Minister.

Although organizational reforms have been the primary lever used to undermine academic freedom, the threats to its individual aspect should not be minimised. Reforms have indeed been accompanied by increased pressure on researchers through State-aligned media, which labels government critics as wasteful or politically motivated, rather than driven by science (Kováts and Rónay, 2021). Fidesz and pro-Orbán media have pressured researchers through campaigns that accuse them of collaborating with NGOs or affiliating with institutions perceived as opponents. These campaigns have been particularly aggressive against academics, often using traditional populist rhetoric targeting elites and minorities, for instance in the campaign against Central European University, directed toward scholars and frequently employing antisemitic rhetoric (Jenne, Bozóki, and Visnovitz, 2022).[8] We provide more details on the political context and the threats to researchers and academic freedom in Appendix 1.A.

III — Academic Data Collection

We conduct our analysis using bibliometric data coming primarily from two types of sources. The first type consists of several national-level databases on research activities specific to each country included in our analysis. The second is OpenAlex, an open-access, world-wide publications database constructed from Microsoft Academic. We combine these sources to build our final datasets.

1. Current Research Information Systems

Our first type of data source is an aggregation of several national Current Research Information System (CRIS) databases. These datasets are national-level repositories that document all research activities within a country, typically established in the 2000s as part of a global effort to enhance research transparency. These datasets generally contain either data about researchers and their publications or project-level data, though some are publication-level databases. In addition to Hungary, we also consider other countries from Central-Eastern Europe, specifically Croatia and Slovenia.[9] All of these databases are part of the EuroCRIS initiative, ensuring they comply with similar standards,[10] and thus adhere to the Common European Research Information Format (CERIF), which allows us to exploit these databases together in spite of their idiosyncrasies. Researchers also have some level of control over the information which appears on their profile and the publications that are listed in the repositories.[11]

For each source, we download data in the available format and apply a translation and harmonization process in order to be able to append all different datasets into one. We then use the data we collect in order to build panel data, at the author-year level, that accounts for the academic performance of each researcher. In addition to publication-level data, CRIS databases collect extensive author-level information, including full names, fields of research,

[8] Jenne, Bozóki, and Visnovitz (2022) give the following reason for this "Sorosozás", which they translate as "Soros-bashing": capitalizing on existing antisemitic tropes to compete against the explicit far-right party Jobbik during the general elections campaign of 2018, including antisemitic rhetoric already directed at Soros from the 1990s emanating from Csurka's MDF against which Orbán pushed back at the time. The authors explain that this campaign was instrumental in helping Fidesz secure a third veto-proof majority in 2018. We further link this battle with the ban on all Hungarian universities (but applying only in practice to Eötvös Loránd University) to offer a master's degree in gender studies, considered to be progressive propaganda pushed by George Soros.

[9] We do not consider Poland, due to the fact that the country was led by the right-wing populist Law and Justice party from 2015 to 2023, again with direct reference to Orbán's governance. We also do not consider Austria due to the relocation of the Central European University to Vienna that could create a bias in our estimation. In addition, we exclude Czechia and Slovakia because their data cannot be used for with author-level analysis.

[10] See `https://eurocris.org/` for more information.

[11] While this gives us confidence in the accuracy of author-level profile information, there could be a selection bias in the publications that indeed appear to be linked to each author. We address this concern with several robustness checks in our results section and in Appendix 1.B.

PhD completion year, and unique identifiers for each author. For authors for whom we have the information, we harmonise the field of research across CRIS data bases.[12] This process results in a large sample of academics from the three countries: in its raw form, the dataset includes information on over 60,000 researchers in Hungary, 40,000 in Slovenia, and 50,000 in Croatia, who entered academia after 1950. To the best of our knowledge, we are the first to use these CRIS data sources in the science of science and innovation economics literature. These data offer broader, more accurate coverage of research activities in smaller countries, without introducing bias related to researchers' fields or productivity that could occur in international bibliometric data bases. We believe these specificities allow us to precisely identify the effect of political pressures on academic performance. Additionally, these national repositories are mainly used by researchers who are citizens of the countries in which they work. We argue that this nationality-based selection is not an issue for our analysis, as "national" researchers still make up the vast majority of academics in all the countries considered. Moreover, political pressure is likely to have a stronger impact on national researchers, while affecting foreign academics in Hungary differently.

2. International Bibliometric Data

The second major data source we use is OpenAlex, an open-access bibliometric database created from Microsoft Academic in 2022. It contains data on 247 million works, 90 million authors, and 60,000 concepts (Priem, Piwowar, and Orr, 2022). We use OpenAlex data for publications from 1990 onward. While Scopus and Web of Science have often been considered more detailed in terms of co-author networks, Microsoft Academic was regarded as one of the most comprehensive databases, particularly for publications not written in English (Visser, van Eck, and Waltman, 2021).

Using both publication-level and author-level data from OpenAlex, and after correcting for some measurement errors[13], we construct a panel of authors active before 2010 in Hungary and in the other Central European countries we study. This database allows us to create an author-year level panel that includes various measures of research quantity and quality for each author, as well as data on their co-author networks and academic affiliations. While national CRIS databases already provide extensive information on individual researchers' activities, OpenAlex is crucial for capturing *international* outcomes. Specifically, it enables us to track researchers' international mobility through their affiliation and their collaborations with foreign researchers who are not included in the national repositories data.[14]

To jointly use these data sources, we proceed as follows. First, we normalize publication titles[15] and collect common external identifiers (Digital Object Identifier (DOI) and Scopus identifiers) for publications across all databases, both national repositories and OpenAlex. We then merge the databases pairwise, matching publications based on identifiers, titles, and publication year, keeping only unique matches. These matched publications are then used to link author profiles across databases using author names. Additionally, we perform a second round of matching based on authors' names, affiliations, and research areas. We treat the national databases as the primary source of truth in this process, as they are curated directly by researchers and/or universities, whereas OpenAlex relies on a clustering algorithm to generate

[12] We provide more details on the collection of the data, its harmonization and the construction of our final panel data set in Appendix 1.B

[13] See Appendix 1.B.3.

[14] In national repositories, names of researchers who do not work in the corresponding country often appear in the list of co-authors. They however do not have any unique author ID, and we thus use OpenAlex in order to be sure they are affiliated with a foreign-based institution.

[15] Specifically, we convert titles to lowercase ASCII characters, remove stop-words, and eliminate non-alphanumeric characters.

author profiles due to its larger scope[16]. Using this method, we for instance match 59,629 author profiles from the Hungarian CRIS to 57,536 OpenAlex profiles and 16,401 profiles from the Slovenian CRIS to 19,432 OpenAlex profiles. We are also able to match 428,472 publications from the Hungarian CRIS and 101,376 from the Slovenian one.

3. Measures of Academic Performance and Mobility

a) Quantity, Quality and Network

The main outcome of interest is the quantity of research that is produced. As the baseline outcome of interest, we use the raw number of publications by a given author in a given year. We then subdivide this number depending on the language of the publication. More precisely, we distinguish the number of publications in English from the number of publications in the "national language" of the country[17].

To account for the quality of research outputs, we use the rank of the journal in which an article is published. Specifically, we compute the average journal rank in OpenAlex using the Scopus CiteScore methodology, which is based on citations[18]. Journals are categorised into the following groups: bottom 50%, top 10%, top 5%, and top 1% cited journals. We then count the number of publications per author, per year, in each journal category. This method is particularly useful since citation counts per publication are not consistently available across all CRIS sources. Additionally, citations are only considered within national repositories, which makes them primarily a proxy for the influence of research within the national research community, rather than an indicator of overall quality[19].

We further measure the impact of the political shock on researchers' collaboration networks. For this, we count the number of collaborations per researcher per year. We define a collaboration as a co-authorship link between two researchers; multiple joint publications within the same year between the same authors are counted as a single collaboration during a given year. We also count the number of new collaborations, defined as first-time co-authorships. For these variables, in OpenAlex, we also distinguish collaborations *within the country* from co-authors that we identify as being from a different country, thus constructing a measure for international collaborations.

b) Research Direction

We also measure research direction by encoding the titles of researchers' articles using a sentence transformer E5 model trained on a multilingual corpus. Similarly to the method of Arts, Melluso, and Veugelers (2025) for all of OpenAlex, we then compute for each researcher present in all CRIS the average cosine similarity between the titles of the articles published each year and the titles of the articles published in the period prior to 2005. We call this measure similarity with past research. We also compute the cosine similarity with the average embedding of political allies and opponents in the current year, either those who work in the same domain (Social Sciences, Health Sciences, etc.), field (Economics, Chemistry, etc.) or subfield (Applied Mathematics, Genetics, etc.) as defined in the OpenAlex classification. We provide more information about these measures in Appendix 1.E.

[16] We give further details on our data analysis and decisions regarding multi-profile authors in OpenAlex in Appendix 1.B.3.

[17] For instance one publication in Hungarian for a Hungarian researcher is counted equivalently to one publication in Slovenian for a Slovenian researcher.

[18] Following James et al. (2019), we compute the journal rank by considering citations received by each journal relative to the total document counts over a four-year window in OpenAlex. This ranking is then merged with each publication in the CRIS data using the ISSN of the publication source. The rank is taken at the field level. We do not use the impact factor provided by OpenAlex due to its limited availability prior to 2012.

[19] Since this variable is only available in the Hungarian national repositories, we mainly use it to assess academic influence among peers in Hungary.

c) Researcher Mobility

Finally, we measure researchers' mobility, considering only migration out of their *home* country, defined as the country in which the researcher is listed in the CRIS data or identified through their affiliation in OpenAlex. OpenAlex provides affiliation data for each year an author publishes, allowing us to track their mobility between affiliations. We create the following indicator variables: whether the author has published in their home country that year; whether the author has published outside their home country that year[20]; whether the author has not published in their home country for a specified number of years; and whether the author has stopped publishing in their home country, meaning they are no longer seen in the home country after that year. Additionally, we account for exits from research altogether, measured by whether a researcher leaves academia, based on the last year they published.

IV — Empirical Strategy

1. Sample Selection

This paper addresses two distinct aspects of the impact of political pressure on innovation, and we therefore construct two sets of samples. The first focuses on identifying the direct effects of political repression on researchers within Hungary. The second concerns the broader impact of illiberal policies on the state of academia in Hungary, thus comparing Hungarian researchers, who were exposed to illiberal policies, and their counterparts in other Central and Eastern European countries.

a) Hungarian Scholars and Political Targets

We first focus on differences between researchers within Hungary based on their perceived political alignment. From the raw sample of researchers in the Hungarian CRIS database, we only keep those with at least one publication between 1990 and 2005, and whose entry year falls between 1990 and 2005. Among these researchers, we identify those who could be perceived as political opponents or allies of the government after the 2010 elections, relying exclusively on authors' individual characteristics measured before 2010.[21]

(i) Political Opponents – This sample aims at understanding the trajectory of researchers that we have identified as perceived to be "political opponents" of the regime. We do so based on two different definitions. First we consider all researchers that are affiliated with the Central European University prior to 2010, as members of the CEU have always been identified by conservative politicians and members of Fidesz as clear opponents of Viktor Orbán. Second, we establish a list of politically involved researchers that collaborate closely with Non-Governmental Organizations that are particularly hostile to Viktor Orbán's policies. In total, the list comprises 440 researchers. All researchers belonging to this political opponents group have been matched to their identifiers both in CRIS data and OpenAlex data by hand[22].

(ii) Politically Targeted Topics – While scientists perceived as political opponents have been directly targeted through laws and media action, this is not the case for the vast majority of researchers. Others may still feel that they are indirectly a target of the regime, particularly due to the topics they focus on. We propose here an attempt to capture how working in a field

[20] For both of these variables, the indicator takes the value 0 if the author has not published that year.

[21] For robustness, we consider alternative time windows to define researchers' alignment, restricting the definition to pre-2005 characteristics.

[22] We provide more details about how these researchers are chosen and how they are matched in all data bases in Appendix 1.C.

of research negatively perceived by the Fidesz governments could expose an author to a chilling effect. We thus focus on researchers publishing on two types of topics. First, those who work on gender-related issues, a subject that has been demonised in the press, and forcibly excluded from the programs of universities. Gender is also a subject which spans more than one domain of science: gender and sexual differences have implications not only in social sciences, but also in health sciences and life sciences. Section 1.C.4 shows that most of the researchers that we identify as working on gender-related topics belong to these two scientific domains. Second, we define a set of authors working on *sensitive topics*, *i.e.*, on topics that are often heavily politicised, dealing with issues that might go against the government's discourse on national history (Deportation, Antisemitism, Roma persecution) or on minorities (refugees, humanitarian crisis, etc.)[23]. We use a list of keywords to understand whether the title of a publication refers to gender and sexual issues or a *sensitive topic*. We consider that researchers with at least three publications or more in these topics between 2000 and 2005 are likely to be specialised. We find in total 1,082 researchers working on gender and 890 researchers working on *sensitive topics*[24].

(iii) Political Allies – This sample focuses on researchers who can be considered politically supportive of the regime. We identify these individuals using the membership list of the Batthyány Society of Professors[25], a conservative academic circle closely aligned with Fidesz, and the website of the Mathias Corvinus Collegium, a think-tank closely linked to the regime[26]. We treat membership in these organizations as indicative of being perceived as an ally to the government. This sample comprises 171 researchers.

In summary, we distinguish four non-overlapping groups within the full raw sample of Hungarian researchers. On one hand, among perceived political opponents, there are academics perceived as political opponents *(i)* and researchers working on targeted topics *(ii)*. On the other hand, researchers aligned with the regime and illiberal policies *(iii)* are hereafter called allies. The remaining researchers belong to the *perceived as neutral researchers* group. This final group is the control group of our different political allies or perceived political opponents. In other words, we compare the research performance and career trajectories of political opponents and conservative researchers to their neutral counterparts, matched based on similar observable characteristics, namely first year of publication, scientific field and vigintile of total publications over the period 2000-2005. Additionally, we ensure that control researchers are not co-authors of either political opponents or conservative researchers, for instance using the network measures that are described in the following subsection. The final samples have the following compositions: 226 political opponents are compared to 498 neutral researchers, 933 researchers working on gender with 10,667 neutral researchers, 832 researchers working on *sensitive topics* with 10,049 neutral researchers, and 58 researchers loyal to the government with 193 neutral researchers.

Network Analysis

All authors from the former samples are part of fully defined, identifiable groups that are politically aligned. The same reason why our methodology allows us to be certain that they are exposed to the political shock also prevents us from assessing the effect of that shock on other researchers who may diverge from the regime or be aligned with it. Those researchers could be heterogeneously affected in a similar manner, but are not part of structured groups. Furthermore, outcomes of the aforementioned well-defined group may have spillovers on connected

[23] Refer to Appendix 1.C.4 for the full list.

[24] We also consider for robustness researchers working on *climate change* issues, who are often targeted by illiberal governments in other contexts, though less so in Hungary

[25] Founded in 1995 for professors "committed to traditional European social virtues". See the following link for more detail: `https://pbk.info.hu/archiv/index_en.htm`.

[26] Not to be confused with the Corvinus University of Budapest, which is a university with no such political alignment.

researchers. In order to capture these potential effects, we make the assumption that researchers are more likely to work with researchers to whom they are politically closer, in particular in contexts in which political affiliation becomes more salient. Using these connections, we build a sample of authors *connected* to the samples defined above.
Not all co-authors have relationships with the same depth and interdependence. We thus must take into account the fact that a co-author of a political opponent on a single paper does not have the same exposure to the political shock as a long-term, frequent collaborator. We use the dependence index from Aghion, Antonin, et al. (2026) to measure this proximity:

$$C_i^g = \sum_{\ell \in \mathcal{P}_i} \omega_{i\ell} \cdot \sigma_\ell^g$$

in which C_i^g is the index of proximity between individual i and group g over the 2000-2010 period. For each publication ℓ in the publications $\mathcal{P}_i$ of author i, we compute $\omega_{i\ell}$, the ratio of citations to publication ℓ to the total citations received by author i for all their publications over the period. We then sum all $\omega_{i\ell}$, weighting them by σ_ℓ^g, the share of authors of publication ℓ that belong to group g, i excluded. In other words, C_i^g measures the share of citations for author i that can be attributed to collaborations with group g, weighted by the intensity of these collaborations.

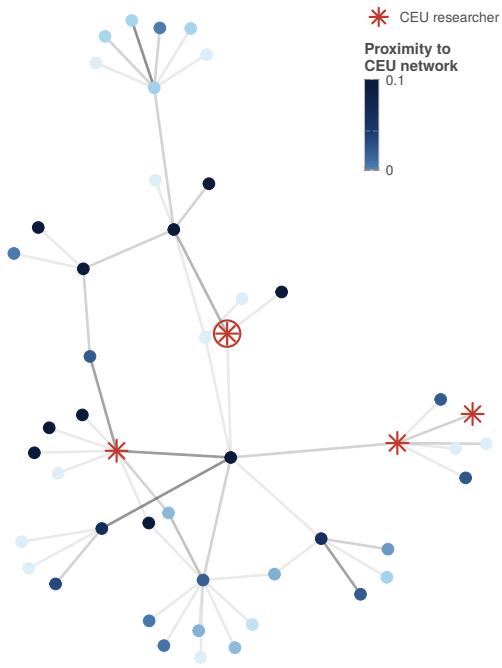
This index only measures the intensity of dependence on the affected groups for direct co-authors. But there is no reason to believe that spillovers or political closeness would stop after the first degree in the network. We therefore expand this definition to also take into account higher degree links between the politically aligned groups and other researchers. Let us define W as the co-authorship matrix, in which $w_{i,j} = \sum_{\ell \in \mathcal{P}_i \cap \mathcal{P}_j} \omega_{i\ell}\omega_{jl}$ measures the strength of collaborations between i and j. We transform this matrix into $\bar{W}$ so that the sum of each row equals 1. Let us also define $\mathbf{c}^g = (C_i^g)_i$ as the vector of direct indices from the first-degree index. We consider that the propagation of the shock across the network decays with a parameter $\lambda = 0.5$, up to the third degree of collaborations[27]. This yields the following continuous index:

$$C_i^{g,\,\text{net}} = C_i^g + \lambda \left(\bar{W}\mathbf{c}^g\right)_i + \lambda^2 \left(\bar{W}^2\mathbf{c}^g\right)_i$$

We provide a simplified example of the co-authorship network in Figure 1.1, for one researcher belonging to the Central European University, represented by a circled red star. Other researchers from the same group $g = CEU$ are also represented with a star, and intensity of the value $C_i^{g\,\text{net}}$ is shown in lighter to darker blue. Some researchers that are only connected to group g through another link may have a stronger $C_i^{g\,\text{net}}$ than some first-degree co-authors because more of their citations come from publications with close co-authors of group g. As the different values of $C_i^{g\,\text{net}}$ on the graph show, this dependence is taken into account by our measure. Throughout our analysis, we refer to a researcher as being part of the network of a given group if their value of $C_i^{g,\,\text{net}}$ is strictly above the median of the value in the dataset for this group (*Opponent Network* or *Ally Network* for instance). After removing researchers who belong to both groups, we keep 5,845 members of the *Opponent Network* and 4,529 members of the *Ally Network*. In order to remove selection bias in our regressions, we compute placebo networks made from groups matched from opponents and allies, respectively yielding networks of 10,749 and 6,483 researchers. We finally remove all authors who belong to two given networks at the same time and match researchers from each network with researchers from the corresponding control network on the same observables as in the above section. The resulting samples comprise 3,592 members of the *Opponent Network* with 2,544 researchers from the control network for

[27] If we considered only first-degrees co-authorships, then $\lambda = 0$ and $C_i^{g,\,\text{net}} = C_i^g$. Results in the following sections are robust to different values of λ.

Figure 1.1 : Opponent Network Example



Notes: The schema above reports an example of the network that we identify when computing the collaboration index. The displayed network is the network of the researcher at the centre of the schema, represented by a circled red star (in that example $g = CEU$, and the schema displays up to 3 degrees of collaborations between authors).

opponents, and 1,899 members of the *Ally Network* with 1,028 researchers from the control network for allies[28].

b) **Hungarian and Central-Eastern European Researchers**

In addition to comparing the trajectories of researchers *within Hungary*, we aim in Section VII to measure the global effect of illiberal policies on academia, performing cross-country author-level comparisons between Hungarian researchers and researchers based in Slovenia and Croatia. Such comparisons between countries are not often performed[29]. We however argue that such comparison is possible here, thanks to the quality of the data we use, which allows us to build a precise matching procedure based on author-level characteristics. We start with the same restrictions as above, in all of CRIS data, only keeping those with at least one publication between 1990 and 2005, and whose entry year falls between 1990 and 2005. This yields in total 36,731 Hungarian authors, 15,728 Slovenian researchers and 21,912 Croatian researchers. To ensure comparability between the treated and control groups, we perform exact matching over the selection period (2000-2005), allowing for replacement, using the raw sample of authors we previously obtained. The matching variables include the year of entry of the researcher in academia, the field she works in and the quantile this researcher belongs to in terms of publication in the CRIS data.

2. Comparing Academic Performance Across Researchers

For each sample we define, we estimate a difference-in-differences regression comparing the group of researchers of interest with a control group drawn from the *researchers perceived*

[28] See Table 1.C.1.
[29] See for instance Kwiek (2016), for a cross-national comparison of highly productive academics in Europe.

as neutral group. We formally investigate the effect of political pressure on researchers' career trajectories by sequentially estimating the following equation:

$$Y_{it} = \sum_{t=2005}^{2024} \beta_t D_i \times Post_t + \alpha_i + \delta_t + \sum_{s=2005}^{2020} X_i \times \mathbb{1}_{\{t=s\}} + \epsilon_{it} \tag{1.1}$$

where Y_{it} is the outcome variable measuring academic performance for researcher i in year t. The treatment indicator D_i takes the value 1 if the researcher belongs to the group of interest and 0 if assigned to the matched control group, as defined in Section IV.1. The variable $Post_t$ is a binary indicator equal to 1 for years after 2010, the onset of illiberal policy implementation. The model includes individual fixed effects (α_i) and year fixed effects (δ_t). The term X_i represents a vector of time-invariant researcher characteristics, interacted with year fixed effects to control for their influence over time. In our baseline specification, X_i comprises only the field and the entry year of the author. We also consider extended specifications that include additional controls such as the author's main domain of research and the number of publications in the first two or five years of a researcher's career.
In all estimation tables, we further decompose the post-2010 effect of illiberal policies into distinct dummy variables corresponding to each mandate of Viktor Orbán (2010–2014, 2014–2018, 2018–2022, post-2022). This allows us to capture not only the overall impact of illiberal policies but also their potential intensification over time, which is likely to be gradual as the political pressure intensifies over the years and mandates.

Each exercise relies on a different treatment group. The first set of results focuses on the "core" opponents and allies, i.e., the individuals identified to be perceived as *allies* or *opponents* in Section IV.1.a. Once the exact matching procedure is performed, the trajectories of these researchers are compared to those of their control group. In addition, we run similar regressions in which treatment is defined by membership in the *Ally Network* or the *Opponent Network*. For these groups, since the collaboration index we build is based on authors' publication and citation counts, we construct collaboration indexes for two additional groups, *ally control* and *opponent control*, taking as the "core" of each network the matched controls of the corresponding treated group. For individuals belonging to the *Ally Network* or the *Opponent Network*, the control group thus consists of the networks of the matched counterparts of the "core" researchers described above. In this way, we compare not only individuals but also networks with similar structure and productivity before the 2010 elections[30].

While the estimating equation remains the same throughout the analysis, we use it to estimate different moments in the data. We believe that the first set of effects estimated in this paper, while the most relevant to our research question, represents additional effects compared to the neutral group, positive or negative. The neutral researchers serve as a control group for the first part of the analysis, but we believe that they are also treated by illiberal policies. They become the treated group in our cross-country comparison in Section VII.

V — How Do Political Pressures Affect Scientists' Productivity?

In this section, we first show that political pressure differentially affects the academic productivity of researchers after 2010, depending on their perceived political alignment, in

[30] Figure 1.C.1 shows the distribution of the collaboration indexes we build for the control group of ally researchers and the control group of opponent researchers.

terms of both quantity and quality (Section V.1). We also show that the effects are similar for researchers targeted through their affiliations or organizations they collaborate with, or through the topics they work on. Second, we investigate the mechanisms through which this negative impact on productivity operates (Section V.2).

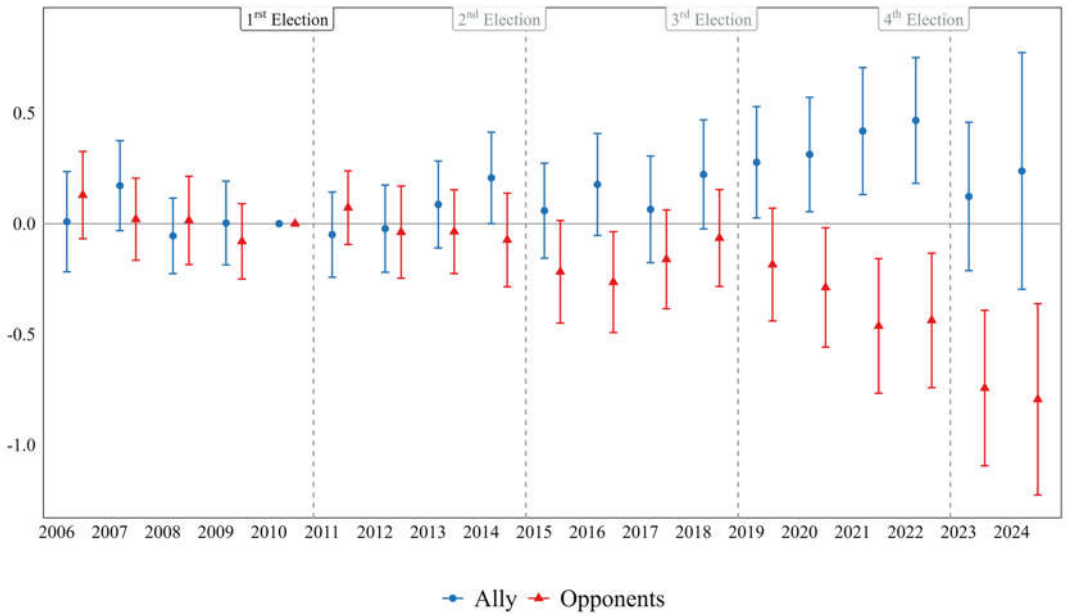
1. The Effect of Being Perceived as an Opponent on Academic Productivity

a) Perceived Political Opponents

This section begins by investigating the effect of illiberal policies on researchers who were directly targeted and perceived as political opponents. Our main results come from estimating Equation 1.1, using perceived political opponents as the treated and researchers from the neutral group as control, matched on pre-2005 characteristics. These first results thus rely only on the "core" of the opponent network, as defined in Section IV.1. We contrast the results obtained on this sample with those obtained on the sample of political allies. Figure 1.2 shows the event-study graph for this first specification, using the yearly count of publications as the measure of productivity. When compared to neutral researchers, political opponents experienced a significant decline in the number of publications. The dashed lines in the graph represent Orbán's successive elections and re-elections. The event-study shows that the magnitude of the effect increases over the later mandates, which suggests that the more intense the political pressure through the years, the stronger the impact on targeted researchers. By contrast, allies' estimated coefficients, if anything, show a slightly positive trend in their total number of publications, especially after Orbán's third consecutive victory in the 2018 elections, although it disappears after 2022. Figure 1.3 reports the average treatment effects estimated for both groups on all quantity and quality outcomes. The negative effect on productivity holds for publications in both English and Hungarian. It represents an average decrease of 18% in the number of publications. Using the value of the pre-2010 average, this represents a decrease of 0.63 publications per year per author. We also document a strong negative effect on research quality: the figure shows the estimated effect on the number of publications in journals ranked in the top 5% and top 10% by citations. The estimate we measure represents a decrease of 59% in the number of publications in top 10%-ranked journals by citations i.e., -0.11 such publications per year per author. This effect is quite important since the pre-2010 average number of such publications per year is 0.19 publication. Researchers perceived as political allies do not appear to be overall affected, while those perceived as political opponents are affected over the different research quantity and quality measures we test.

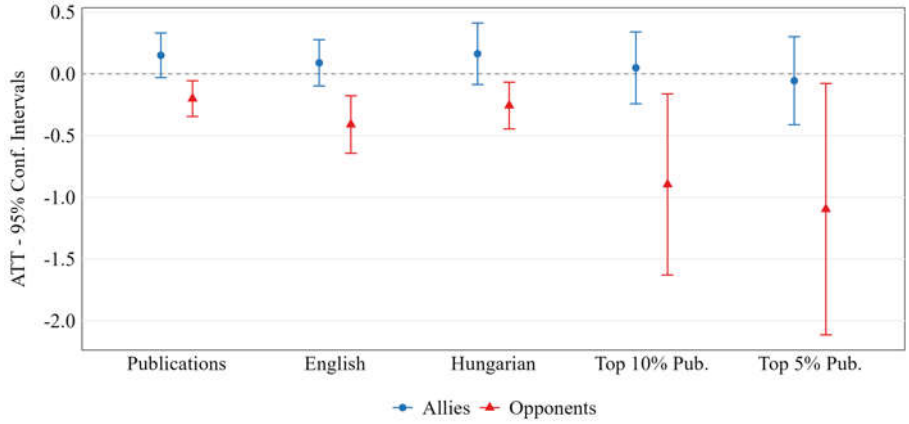
Second, we do not restrict ourselves to the "core" of both perceived political opponents or allies, but also demonstrate that these effects are propagated along their network. Table 1.F.5 shows smaller effects of the shock on researchers who are part of the network of opponents on the total number of publications and on the total number of publications in high-quality sources. The effect only becomes significant for the second and third mandates. When statistically significant, the effect that we measure for the network of political opponents is of a smaller magnitude than the one we measure for the individual at the core of the network, indicating that there is attenuation compared to perceived political opponents: it represents for instance a decrease of 8% in total publications during the second mandate, and a change of -22% in the count of publications in the top-10%-ranked journals. In the following section investigating the mechanisms of this effect, we show that indeed researchers at the centre of the networks of political opponents are more impacted than the individuals located at its periphery. We also summarize the results for ally researchers in Table 1.F.1. These researchers do not incur any loss in terms of academic productivity, neither in the quantity nor in the quality of their research.

Figure 1.2 : Effect on Publications, by Political Alignment.



Notes: The graph above reports the coefficient in Fixed-Effects PPML regressions on publications for each political group compared to Hungarian researchers that are not identified as being close to any organization targeted by the government nor to researchers explicitly supporting the government. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable.

Figure 1.3 : Effect on Hungarian Academics' Productivity, by Political Alignment.



Notes: The graph above reports the ATT coefficients in Fixed-Effects PPML regressions for each political group compared to Hungarian researchers that are not identified of being close to any organization targeted by the government. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable.

b) **Targeted Topics**

We then turn to the analysis on the effect on researchers perceived as political opponents because of the topic they work on. We estimate equation 1.1 again, using as a treatment whether the author worked in gender-related topics or sensitive topics, i.e., if they had at least three publications featuring the associated keywords prior to 2005. Table 1.1 summarises this effect on researchers' productivity. We find that both for gender-related and sensitive topics their productivity is negatively impacted, both in quantity and quality, similar to our results for political opponents. The magnitude that we measure is similar to the previous section: gender-topic researchers lose roughly 13% of total output after the policy change, with a fairly uniform hit across languages and a somewhat larger drop of 18% at the top of the quality distribution. Sensitive-topic authors show a similar pattern but with a higher effect on publication in top 5% ranked journals in terms of citations, as it represents a decrease of 25%. Targeted topics are thus affected substantially, although the magnitude of the effect we find is smaller than the one on academics who are directly perceived as political opponents, but of a higher magnitude compared to their network[31].

These findings — although limited to researchers working on gender-related or "sensitive" topics[32] — are for instance consistent with the qualitative evidence presented by Kováts and Rónay (2021), who document how concerns about potential government targeting can have tangible effects on researchers' output.

As we perform the exact same analysis as in the previous section, we only focus on researchers who entered academia in 2005 or before. The negative impact that we document on these researchers is, however, likely to also exist on researchers who entered after this year, especially after 2010. This assumption is supported at least descriptively, as Figure 1.C.7 shows. The figure plots the number of new researchers each year who publish more than once in gender-related topics in their first five years of activity. This count starts declining from 2015 on, which suggests indeed that the topics related to gender studies become less attractive to young researchers over the period, which likely explains part of the negative productivity effect we find.

For all the regressions presented in this paper, we vary the variables on which we match the control to the treated group. Our preferred matching equation relies on variables that capture essential characteristics of researchers when it comes to comparing their trajectories both across-country and within Hungary. However, we also create alternative samples, using the same exact matching procedure but relaxing it by dropping some of the matching variables. Doing so, we logically increase the size of both the control and treatment groups. We find that our results are robust to this exercise, and that samples with increased size show similar effects of illiberal policies on targeted researchers. We perform another series of robustness checks regarding the results about gender studies. First, we choose more restrictive conditions to select authors working in gender topics and find similar results. Second, we replicate these results in OpenAlex, in which we find that the overall number of publications and average quality decline as well. Results are also robust to different winsorization thresholds on the outcome variables (0.1%, 1% and 5%), showing that our results are not driven by extreme values in the distribution of outcomes, a usual concern when using publication data.

[31] We also find that there appears to be a stronger effect for researchers cumulating both statuses, and discuss the results in Section 1.F.4.

[32] In particular, they cannot be extended to researchers working on topics related to climate change, as shown in Table 1.F.7.

Table 1.1 : Effect on Academic Productivity - Targeted Topics

	Publications		Publications (EN)		Publications (HU)		Pub. Top-5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
gender × post	-0.143***		-0.170***		-0.133***		-0.197***	
	(0.025)		(0.031)		(0.035)		(0.044)	
sensitive × post		-0.125***		-0.101***		-0.163***		-0.288***
		(0.025)		(0.039)		(0.038)		(0.111)
Mean dep. var. (pre)	3.398	4.694	1.364	1.364	2.497	2.497	0.528	0.381
Observations	202,535	190,362	168,138	145,064	190,507	181,489	215,173	46,174
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those who published at least two publications in gender topics for the first line and sensitive topics for the second line. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total publications (columns 1-2), publications in English (columns 3-4), publications in Hungarian (columns 5-6) and publications in the top 5% most cited journals in OpenAlex. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

2. Mechanisms of Repression

This section aims at understanding how the political pressures translate into worse or better research outcomes for the affected groups. We consider different potential channels. First, we investigate the funding channel, and find no evidence that funding is the main margin through which treated researchers experience a change in their productivity. Second, we consider their network of collaborations and find that link formation appears to have been disrupted for political opponents, likely explaining the effect on the quantity and quality of research. Finally, we consider an effect closer to proper silencing: a loss or gain of influence due to implicit incentives to cite the work of aligned researchers differently also appears to explain part of the productivity effect.

a) Is it Just Funding?

We first show that while funding could play a major role for the country as a whole, it does not seem to be the main channel through which targeted researchers were affected. A very efficient and direct way to conduct political repression would be to diminish the budget allocated to opponents, at the institutional or individual level. One observation that could be first made is that the core of the individuals we identify as political opponents are identified through their affiliation with privately funded institutions, such as the Central European University, which is funded by a private foundation independently from the Hungarian State. As we have no access to funding data at the institutional or individual level, we use several indirect tests that exploit differences in researchers' dependence on public funding. Funding at the country-field level reveals no particular downward trend over the entire period for any large research area.[33]

First, we leverage the difference in the public or private status of researchers' affiliation. We subdivide the control and treated groups based on whether they had been affiliated to a public institution before 2010. If political pressure operated mainly through public funding, political opponents affiliated to public institutions should be more penalised than other opponents. Just like the above regressions, the triple difference-in-differences is estimated on a matched sample

[33] We plot data from the OECD Government Budget Allocation for Research and Development in Figure 1.A.5 and find no particular targeting.

with the exact same matching variables. We estimate the following extension of equation (1.1):

$$Y_{it} = \beta_1 D_i \times Z_i \times Post_t + \beta_2 Z_i \times Post_t + \beta_3 D_i \times Post_t + \alpha_i + \delta_t + \sum_{s=2005}^{2024} X_i \times \mathbb{1}_{\{t=s\}} + \epsilon_{it}, \quad (1.2)$$

where D_i takes value 1 when we identified the researcher as a political opponent and 0 for matched neutral researchers, and $Post_t$ takes value 1 for years after 2010. Z_i is an indicator for the dimension of the funding heterogeneity we study: affiliation to a public institution before 2010. As in equation (1.1), α_i and δ_t are individual and year fixed effects, and X_i represents field and entry year of the author, interacted with year fixed effects. The term $D_i \times Z_i$ is absorbed by individual fixed effects, and the coefficient of interest, β_1, measures the additional effect of political pressure on opponents for whom $Z_i = 1$.

Table 1.2 presents the results. The penalty for opponents is robust to controlling for being affiliated to a public institution: relative to matched neutral researchers, they publish 12% fewer papers after 2010, 16% fewer in English and 13% fewer in Hungarian. They also receive 38% fewer citations from other Hungarian researchers. The triple interaction, however, is small and not significant for any outcome: opponents affiliated to public institutions are not penalised more than other opponents. These estimates are imprecise and do not rule out moderate differences, but their signs do not point to a larger penalty for publicly funded opponents. We run an additional placebo exercise to show that public affiliation alone is not associated with a post-2010 break: among researchers who are neither opponents, allies, nor co-authors of either group, the interaction between public affiliation and the post-2010 period is small and not significant for any outcome (see Table 1.F.12).

Second, we run a similar exercise using European funding pre-2010, through the public data from the European Research Council on grant winners[34]. ERC grant recipients obtained additional funding independently of the Hungarian government during the period. Only 83 Hungarian researchers have received an ERC grant over the period 2007 to 2025, and in our sample of political opponents, this is the case for only 6 researchers.[35] In order not to take into account the effect of being the researcher who has been granted the ERC, we analyse the effect of exposure to this type of funding by constructing the same collaboration index C_i^g, using all recipients of ERC grants prior to 2010. Similarly to other analyses based on network proximity, we consider a researcher to be exposed to ERC funding when this value is above the median value of the index.[36] We then estimate equation (1.2) with this exposure indicator.

Table 1.F.10 presents the results. They likely provide evidence that even for researchers who are close members of the network of very renowned researchers who received funding from outside sources, this access has not been enough to offset the effect of illiberalism. Exposure to ERC grantees is associated with higher output after 2010 for neutral researchers, with 25% more publications, but not for opponents: the triple interaction is not significant for publications, and it is negative for citations, with a 17% decline. When this exposure index also includes grants awarded after 2010, exposed opponents partly recover, in particular in English-language publications (Table 1.F.11). Since these later grants are themselves an outcome of post-2010 productivity, we only interpret this result as descriptive. Researchers who depend on other sources of funding than the government appear to still suffer from the political climate. Funding itself, while probably playing a role for the majority of researchers, is not the only mechanism through which the illiberal penalty affects researchers.

[34] We use this measure also as a proxy of the access to non-governmental funding. These researchers are indeed also more likely to be able to advertise their work in order to obtain other funding.

[35] Appendix 1.F.5.a presents descriptive statistics on the number of ERC recipients in the period and unconditional trends.

[36] We however consider in this case only researchers who are direct co-authors of the ERC recipient, thus not using the network version of the collaboration index that we describe in Section IV.1.a

Table 1.2 : Trends in Outcomes for Political Opponents, by Public Affiliation

	publications (1)	pub. in english (2)	pub. nat. language (3)	nr co-authors (4)	nr new co-authors (5)
pol. opp × post	-0.7359***	-0.4499***	-0.1680	-0.4972*	-0.4342*
	(0.2086)	(0.1445)	(0.1521)	(0.2808)	(0.2415)
public × post	0.5187***	0.2634***	0.3517*	0.4002**	0.1297
	(0.1968)	(0.0751)	(0.1921)	(0.1847)	(0.1390)
public × pol. opp × post	-0.8819	-0.1046	-0.9060**	0.5726	0.5001
	(0.8888)	(0.6435)	(0.4423)	(0.5579)	(0.4250)
Observations	643,834	643,834	643,834	643,834	643,834
R^2	0.562	0.591	0.496	0.582	0.362
Within R^2	0.000	0.000	0.000	0.000	0.000
Mean.Dep.Var.Pre	2.880	1.325	1.835	3.124	1.827
author fixed effects	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect PPML regressions of equation (1.2), for each outcome, on data from the Hungarian research repository. The unit of observation is author by year and the sample period is from 2005-2024. Treated authors are authors from our sample of political opponents, split by whether the author was affiliated to a public institution before 2010. Control authors are neutral Hungarian researchers matched to the treated, split by whether the author was affiliated to a public institution before 2010. We report the effect on total publications, publications in English, publications in Hungarian, citations count, total number of collaborations and number of new collaborations. Control variables account for trends for field and entry year in academia FE. *** p<0.01, ** p<0.05, * p<0.1.

b) Network Shrinking Effects

We now turn to the analysis of network dynamics for treated researchers compared to the control group. If being the target of political pressures prevents researchers from forming new connections or exploiting existing ones, it can lead to the productivity effect we observe. In particular, after 2018, treated researchers were sometimes named in the press. Other researchers could thus fear that they would become exposed to pressures themselves if they collaborate with political opponents.

In these regressions, we compute the effect of political alignment only on the number of co-authors who are also present in the national database. We believe that these diverging patterns in the Hungarian network of allies and opponents help explain why publications in Hungarian are more affected in both directions than those in English. We perform a similar regression in OpenAlex, on the subset of authors whom we identify from the CRIS datasets. Using the same matching algorithm, we compute the same regression, dividing publications by whether they have any co-author outside of Hungary, a co-author in the region outside of Hungary[37], and a co-author in the US. Table 1.F.13 shows that there is a positive effect building up over time for international collaborations, in particular with co-authors from the US. Conversely, for allies, Table 1.F.14 shows that international collaborations tend to decrease. Putting those regressions together, we find that the network of opponents shrinks in Hungary but collaborations intensify abroad, while the opposite happens for allies.
Table 1.3 shows the results on the political opponents in terms of total collaborations and new collaborations. Researchers perceived as political opponents see their number of co-authors fall by 25% and their number of new collaborations by 28.6% after treatment. The allies' network, in contrast, does not seem to be affected, either in total number of co-authors or of new collaborations. Table 1.4 reports the same outcomes when the treatment group is the whole network of political opponents i.e., individuals whose collaboration index with perceived opponents is above the median. We find similar results, for both the total number of

[37] In Slovenia, Slovakia and Czechia, other Central European countries that entered the European Union at the same time as Hungary.

collaborations and the number of new collaboration, although they are of smaller magnitude compared to researchers representing the "core" of political opponents. Authors with above-median collaboration with perceived political opponents have about 16% fewer co-authors per year on average, which corresponds to about 1 fewer co-author per year. The decomposition per mandates again shows that this effect intensifies over the years, with stronger effect for latest mandates. This attenuation from the core to the broader network of perceived political opponents is consistent with Figure 1.C.3, which shows that authors at the centre of the opponents' network (i.e., with a higher collaboration index) are more affected in their number of collaborations than those at the periphery.

Table 1.3 : Allies and Opponents: Effect on Network and Influence

	Nb. co-authors		Nb. new collaborations		Repeat collab. rate		Nb. citations	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Opponents × Post	-0.288**		-0.337***		0.049		-0.195	
	(0.134)		(0.124)		(0.034)		(0.128)	
Allies × Post		0.041		0.023		-0.026		0.030
		(0.107)		(0.123)		(0.024)		(0.153)
Mean.Dep.Var.Pre	1.47	13.2	1.14	6.01	0.214	0.517	11.7	91.1
Observations	9,942	4,466	9,743	4,413	3,987	3,334	8,818	4,334
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-Entry year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-Field fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those perceived as political opponents (first line) and as political allies (second line). Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total number of co-authors (columns 1-2), total number of new co-authors (columns 3-4), number of co-authors with whom the researcher has already collaborated before (columns 5-6) and citations from other MTMT publications (columns 7-8). The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

c) Loss of Influence

While network reallocation may explain part of the effect, we also find that the influence of political opponents on Hungarian research decreases during the post-2010 period. In the Hungarian national repository data, we have access to the number of citations received by a publication from other publications of the database. We measure quality through the citations received globally by the journal or venue of publication. Citations from the rest of the world would be a measure of quality as well, but we consider that citations within Hungary also account for a different component, influence within the country. If all things equal, a researcher is less likely to cite a publication because it comes from a disregarded group, or more likely to cite it because one of the authors is favoured by the government, then the resulting difference in citations is not due to a difference in quality but rather in how influential the researcher is in the country.

Figure 1.F.5 shows that there is a significant negative effect in an OLS specification of Equation 1.1, reaching around -8 citations received by members of the network of perceived political opponents per year over the period, and seems to decrease right after 2010. However, the effect is no longer significant for the entire period when using a PPML specification, as shown by Table 1.4; the effect is still increasing over time but no longer significant for most periods. However, we find the same attenuation pattern from the core to the broader network as with the effect on the number of collaboration, as reported in Figure 1.C.3. Authors with a higher collaboration index are more negatively impacted than those at the periphery of the network in terms of influence within Hungary. Conversely, Table 1.3 and Table 1.F.15 show

Table 1.4 : Collaboration and Influence - Opponents Network

	Nr Coauthors		Nr New Collaborations		Citations	
	(1)	(2)	(3)	(4)	(5)	(6)
Network pol. opp. × post	-0.176***		-0.164***		-0.060	
	(0.036)		(0.041)		(0.045)	
Network pol. opp. × post (1rst m)		-0.120***		-0.144***		-0.064
		(0.030)		(0.039)		(0.042)
Network pol. opp. × post (2nd m)		-0.204***		-0.175***		-0.108**
		(0.038)		(0.044)		(0.051)
Network pol. opp. × post (3rd m)		-0.223***		-0.197***		-0.110*
		(0.047)		(0.051)		(0.066)
Network pol. opp. × post (4th m)		-0.237***		-0.225***		-0.055
		(0.060)		(0.075)		(0.106)
Mean dep. var. (pre)	6.216	6.216	3.271	3.271	38.217	38.217
Observations	75,087	75,087	74,579	74,579	72,738	72,738
author fixed effects	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as political opponents is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total number of co-authors (columns 1-2), total number of new co-authors (columns 3-4), and citations from other MTMT publications (columns 5-6). The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Odd columns report the average effect on the period 2010-2024 and even columns split the effect by mandate. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

that both for the ally researchers and for their network, we do not find a positive effect mirroring the decrease in citations for members of the network of perceived political opponents. The channel of influence over science only appears to be relevant for political opponents: researchers in Hungary are not more likely to cite political allies of the government after the shock.

So far, researchers have only been considered purely as objects and not as agents for all of those three potential channels. Yet, we believe that scientists who are subjected to political pressures may change their behaviour in order to respond to them. These actions would affect the productivity of researchers in discernible ways, but they would also incur additional costs of reacting to the shock. It is therefore necessary in order to fully understand the effect of political threat to study these individual reactions.

VI — Academic Strategies under Political Pressure

We now present evidence of different behavioural strategies among targeted researchers. We draw on the *Exit, Voice, Loyalty* framework from Hirschman (1972), for which our context provides a suitable empirical test: when in disagreement with an organisation, an individual may choose among different courses of action. In our setting, academics facing political threats and increasing pressure of an illiberal government may react in different ways. We therefore test empirically for three distinct strategies. First, a researcher may *exit*, i.e., quit her position or leave the country. Second, she may *voice* her dissent. Finally, she may remain *loyal* and comply with the new research requirements. We nuance this third strategy and show that, rather than becoming fully silent, researchers are likely to self-censor.

1. Theoretical Framework

We formalise the responses of researchers in a simple static model, presented in Appendix 1.D. Researchers differ in their perceived political alignment with the government, θ_i, and in their returns to migration, m_i, and choose among four mutually exclusive strategies:

- *Exit* consists of leaving Hungary or academia, at a fixed cost, against the returns to migration m_i.
- *Voice* is an active and visible response: the researcher expresses dissent against the government or the illiberal policies, including through her publications. This strategy implies a fixed cost and a cost increasing with the political pressure. Its returns increase with the research visibility and outside option.
- *Loyalty* is the passive, default strategy: the researcher stays and takes no action, choosing to bear the pressure. This strategy exposes to political pressure in proportion to her dissent with government positions. Loyalty in our context is not an attachment that delays exit, but the absence of any response.
- *Self-censorship* is an active but invisible response: the researcher stays and shifts her research away from politically sensitive content, at a fixed cost, which removes the disutility of disagreement.

Self-censorship and exit are two forms of *silence* between which the choice depends on mobility alone: the immobile self-censor and the mobile leave. The model yields predictions that we take to the data (Section 1.D.3): (i) exit has a political component, so it rises with pressure among perceived opponents but not among aligned ones with the same outside option; (ii) political exit is concentrated among researchers with high returns to migration, while the less mobile self-censor; (iii) self-censorship rises with pressure; (iv) voicing is an intermediate strategy, so that dissent should precede exit along the pressure path; (v) all effects fade with distance from the core of the opponent network.

Our empirical strategy follows directly. As in Section V, we compare the trajectories of researchers who entered academia before the shock, perceived opponents or not, using successive mandates as increasing levels of pressure. We then exploit heterogeneity in returns to migration, along three dimensions: field of research, which governs the international transferability of skills; career stage, younger researchers having a longer horizon over which to recoup the cost of leaving; and a direct proxy of individual mobility. Return to migration is defined in our context taking into account three dimensions: first the distribution in terms of their academic outcomes pre-2010 (publications, citations, etc.), second the age of the researchers, and third the domain she works on.

2. Exiting Hungary

We first assess how frequently targeted researchers resort to the "Exit" strategy. In our context, we consider two types of observable outcomes as exit. The first is leaving academia. As we do not observe the actions of researchers outside of publishing, we consider that a researcher "retires" when we do not observe her publishing after a given year. The second is leaving the country. Similarly, while we do not observe the actual migration decisions and their timing, we have access to researchers' affiliations for most publications in OpenAlex. Based on this information, we infer whether they have an affiliation in the country, outside the country or both. Both outcomes are relevant beyond our understanding of reactions of individuals to illiberal policies. These actions have a direct effect on the total production of the country's research and innovation ecosystem, and can thus be interpreted as a loss in human capital. Again, we do not restrict our analysis to researchers who are directly perceived as political opponents and include in our regressions researchers who belong to their network as defined in Section IV.1.a.

This allows us to draw a more realistic picture of the effect of political pressures and how they propagate, while increasing our statistical power.

We estimate again Equation 1.1 as a linear probability model, with variables measuring exit as the outcome. We present the results of these regressions in Table 1.5. We first find that there is no increase in the probability of leaving research altogether: this does not appear to be a margin through which targeted researchers respond. Figure 1.4a shows that descriptively, we observe a pattern of increased exits from the country among members of the network of political opponents, measured by the probability of having an affiliation abroad for a period of four years. This holds true in the difference-in-differences specification: Figure 1.4b shows that on average, this probability to exit increases by about 5 percentage points at the end of the period. For shorter periods of continuous presence abroad, this effect becomes significant as soon as the second mandate of Orbán, prior to the ban of the Central European University, suggesting that we do not only capture the effect of government policy, but also of actions chosen by researchers. The effect is also very strong, reaching over 50% of the pre-period average at the second mandate only. It rises starkly during the third mandate, after the ban: we believe that this increase, which remains stable over the last period, reflects the expulsion of the CEU, which has further incentivised researchers to leave.

In order to validate our findings, we run the same regressions using as a treatment whether the researcher is part of the network of political allies of the government. If what we are observing in the aforementioned regressions is mechanical and due to our selection process - that is, if what we observe is researchers connected to internationally connected researchers leaving more often – we should also find a positive effect on exit outcomes. If what we have shown above indeed represents a strategy of reacting against political pressure, we should find no effect on the network of ally researchers. In Table 1.G.2, we show that there is no significant effect on migration or retirement decisions for members of the network of ally researchers. We interpret this finding as additional evidence that we are accurately capturing strategies of politically targeted researchers.

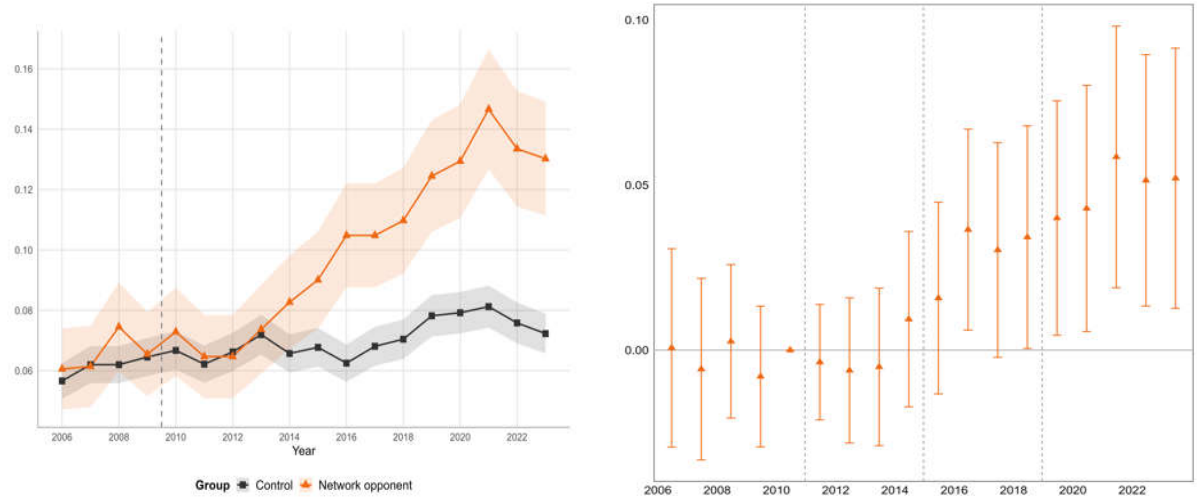
The timing of these effects is in line with prediction (iv): exit becomes significant from the second mandate onward, whereas voicing starts as early as the first years after 2010 (see below). We then test whether exit is driven by returns to migration, as predicted by the model. Panel (a) of Figure 1.5 splits the effect by field. The increase in the probability of leaving among the network of political opponents is concentrated in life and health sciences, where skills are most easily transferable abroad, at around 11–12 pp. It is smaller in social sciences, where human capital is more country- and language-specific, and absent in physical sciences. For allies, the estimates are positive only in life sciences and negative in physical sciences, consistent with the model's prediction that pressure, if anything, retains aligned researchers at the margin.

Panel (b) of Figure 1.5 splits the sample by returns to migration. The effect is entirely driven by researchers with high returns to migration, whose probability of leaving increases by about 4.5 pp, while it is small and non-significant for low and intermediate levels. Among allies, exit does not respond at any level of mobility. Since high-mobility allies and opponents face the same outside options, this contrast isolates the political component of exit (prediction (i)), and its concentration among the most mobile is consistent with prediction (ii): perceived opponents who cannot leave must instead resort to another strategy.

3. Voicing Dissent

The struggle between the Central European University and the Hungarian government was particularly visible, and researchers from the CEU have been outspoken in the media against

Figure 1.4 : Effect on the Probability of Leaving Hungary

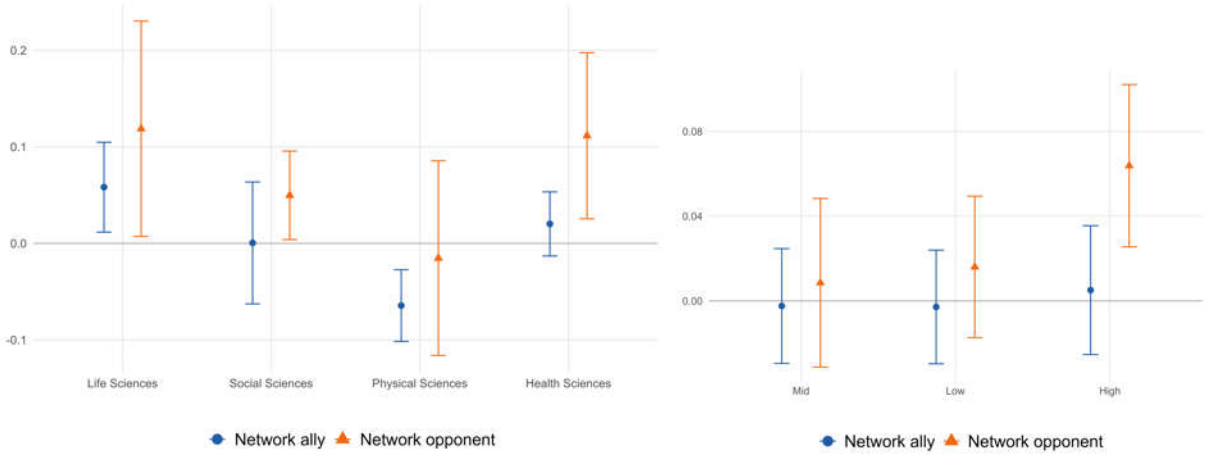


(a) Groups trends

(b) Difference in Probability

Note: Panel a) represents the difference in trends in probability of being away from Hungary over four continuous years in OpenAlex data, between researchers whose collaboration index with researchers perceived as political opponents is above median, and control authors. The latter are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. Panel b) shows the results from Fixed-Effect OLS regressions of the same variable. The unit of observation is author by year and the sample period is 2005–2024. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author.

Figure 1.5 : Effect on the Probability of Leaving Hungary, by Field and by Returns to Migration



(a) By field

(b) By returns to migration

Notes: The graph reports the ATT coefficients from Fixed-Effect OLS regressions of the probability of being away from Hungary over four continuous years in OpenAlex data, with 95% confidence intervals. Panel (a) estimates the effect separately by domain of research. Panel (b) estimates it separately by tercile of returns to migration (Low, Mid, High), as defined by the total number of citations pre-2010 that one researcher has. Treated authors are those whose collaboration index with researchers perceived as political opponents or allies is above median. Authors are matched using their number of publications over the period 2000–2005, entry year in academia and field of research. The unit of observation is author by year and the sample period is 2005–2022. Standard errors are clustered by author.

Table 1.5 : Political Opponents Network: Migration and International Collaboration

	Migration				Collaboration					
	Any affil. abroad		Abroad (2y)		Intl. collab.		Regional collab.		Retiring	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Network pol. opp. × post	0.051***		0.030***		-0.115		-0.017		-0.0002	
	(0.019)		(0.011)		(0.131)		(0.062)		(0.0009)	
Network pol. opp. × post (1rst m)		0.017		0.0002		-0.331**		-0.056		0.0005
		(0.019)		(0.009)		(0.131)		(0.052)		(0.001)
Network pol. opp. × post (2nd m)		0.067***		0.032**		-0.037		0.030		0.0001
		(0.021)		(0.013)		(0.151)		(0.069)		(0.001)
Network pol. opp. × post (3rd m)		0.075***		0.051***		-0.042		-0.026		-0.0009
		(0.026)		(0.017)		(0.167)		(0.083)		(0.001)
Network pol. opp. × post (4th m)		0.050		0.057***		0.015		-0.053		-0.0009
		(0.035)		(0.019)		(0.188)		(0.124)		(0.002)
Mean dep. var. (pre)	0.156	0.156	0.063	0.063	0.313	0.313	2.146	2.146	0.360	0.360
Observations	71,528	71,528	128,772	128,772	69,048	69,048	123,797	123,797	128,772	128,772
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions on data from OpenAlex. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as political opponents is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. First panel reports the effect on migration: probability of having a new affiliation outside of Hungary (columns 1-2), the probability of publishing abroad over 2 years (columns 3-4). Second panel, reports the effect on number of publications with researchers outside Hungary. Last column shows the probability of retiring i.e., not publishing any more. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

this decision[38]. More generally, voicing could be undertaken via social media posts, protests or interviews with traditional media. While academic publishing could be rarer, it is still a common practice in social science, and represents a large, costly effort toward voicing. Researchers in social science may redirect their research in order to study a particular aspect of the regime and use their research to voice directly, which could be interpreted as a "lower bound" effect on all types of voicing behaviour.

In this section, we thus provide evidence that researchers subjected to political pressures have shifted research direction in order to speak out against the government. We estimate our main equation (Equation 1.1) again, but on an indicator variable of publication per author of four different subsets of publications which we consider as indicative of voicing activity[39]. The first two are the gender topics and sensitive topics defined earlier. We add to them two narrower types of topics:

- *Rule of law*: One of the key elements of Orbán's illiberal policies was disregard of the rule of law. This strategy was both openly embraced by the government and noted by its critics, including the European Commission. Emphasis on the rule of law and illiberalism from political opponents can therefore be considered as direct voicing, as it targets one of the salient political features of the government.
- *Minorities*: Another salient element of illiberal policies in Hungary consisted of targeting minority rights or attacking them in the media. Examples include the passing of a law similar to Russia's anti-"homosexual propaganda" law, or the vilification of Soros reactivating antisemitic stereotypes. We consider that publications explicitly mentioning minority rights are likely to be opposed to the government.

[38] For instance, the New York Times published a widely read article in 2018 including a statement from the president of the university (Santora, 2018).

[39] In this section, due to the small number of researchers who publish in any of these subsets, all results are estimated using two-way fixed effects OLS rather than PPML.

We detail in Appendix 1.C.4 how we define these variables from the vocabulary used in the publication title. Members of the network of political opponents strongly increase their total number of publications in which the lexicon of the "rule of law" appears, as early as during the first years post 2010, as shown in Figure 1.6. Again, the magnitude of the positive effect we find increases over the years, suggesting that the higher the political pressure, the more likely academics perceived as political opponents are to express their dissent. We also note this trend does not diminish after 2018, suggesting that these researchers have largely chosen to "voice" their dissent rather than adopt a strategy of "loyalty". Additionally, we show in Table 1.G.1 that this behaviour is not only driven by the Central European University members alone, nor by researchers working with NGOs, but that both publish more on democratic backsliding. In that respect, our results indicate that at first glance, political pressure appears relatively ineffective at fully silencing those perceived political opponents. Conversely, ally researchers and their network do not show similar results[40].

Table 1.6 : Political Opponents Network: Voicing Behaviours

	Rule of law (EN) (1)	Rule of law (HU) (2)	Minorities (EN) (3)	Minorities (HU) (4)	Sensitive (EN) (5)	Sensitive (HU) (6)
Network pol. opp. × post	0.0174***	-0.0040	0.0044***	-0.0103***	0.0185***	-0.0306***
	(0.0032)	(0.0033)	(0.0013)	(0.0034)	(0.0040)	(0.0055)
Observations	226,224	226,224	226,224	226,224	226,224	226,224
R^2	0.303	0.319	0.215	0.302	0.286	0.288
Within R^2	0.001	0.000	0.000	0.000	0.000	0.000
Pre.Mean.Dep.Pre	0.011	0.034	0.004	0.041	0.040	0.112
Author FE	✓	✓	✓	✓	✓	✓
Year × Field FE	✓	✓	✓	✓	✓	✓
Year × Entry year FE	✓	✓	✓	✓	✓	✓

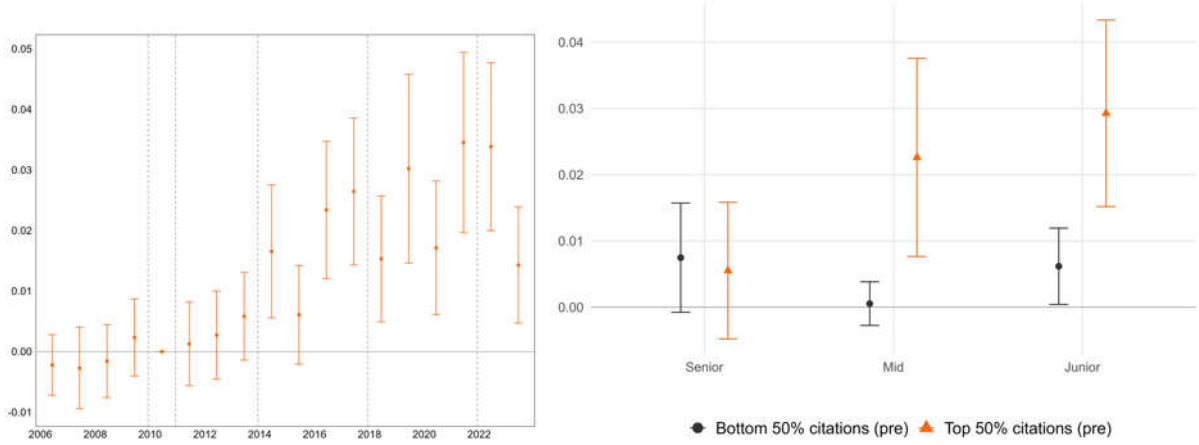
Note: Results are from Fixed-Effect OLS regressions on CRIS data. The unit of observation is author by year. Treated authors are those belonging to the network of political opponents, as defined in Section IV.1. The dependent variables are indicators for publishing on rule-of-law topics, minority-related topics and sensitive topics, separately for English-language (EN) and Hungarian-language (HU) publications. All specifications include author, year × field and year × entry-year fixed effects. Standard errors are clustered at the author level and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

We also estimate the same equation as above, splitting further publications by language. Table 1.6 shows that voicing only occurs in English. For minority-related and sensitive topic publications, the effect actually becomes negative in Hungarian. This recontextualises the results described above: while researchers are more likely to voice overall, this is only due to the fact that they are doing so in publications visible internationally, while instead refraining from voicing in publications within the country. To some degree, these results show that while researchers voice their dissent to international audiences, they seem to be effectively silenced within Hungary. Our interpretation is that the behavioural responses of researchers are interconnected: those expressing dissent against illiberal policies may choose to do so in ways that do not attract negative attention of national authorities, as publishing in Hungarian could. We further investigate this phenomenon in the next section.

Panel (b) of Figure 1.6 splits the effect by career stage and by citations before 2010. The increase in *rule of law* publications is driven by researchers in the top half of the pre-period citation distribution, among mid-career and junior researchers. For these groups, the number of *rule of law* publications increases by about 0.023 and 0.029 per year respectively, both significantly positive. Among senior researchers, the effect is small and non-significant regardless of citations. Among less-cited researchers, the estimates are small at every career stage

[40] See Appendix 1.G for more details.

Figure 1.6 : Effect on the Number of *Rule of Law* Publications, in English



(a) Average effect

(b) By seniority and pre-period citations

Notes: Panel (a) reports Fixed-Effect OLS regression estimates for the difference in the yearly number of *rule of law* publications in English between researchers we identify as political opponents and other Hungarian researchers, over the period 2006–2023. Panel (b) reports the average post-2010 effect estimated separately by career stage (Senior, Mid, Junior) and by citations before 2010 (below or above the median), with 95% confidence intervals. Authors are matched using their number of publications over the period 2000–2005, their main field of publication and their year of entry in academia.

and only marginally different from zero for juniors. Voice is therefore concentrated among the most visible researchers, whose reputation provides some protection and good outside options, and among younger cohorts, whose careers are less tied to the domestic institutions targeted by the government. This pattern is consistent with the model's prediction that voicing is a strategy chosen by perceived opponents who can bear its costs.

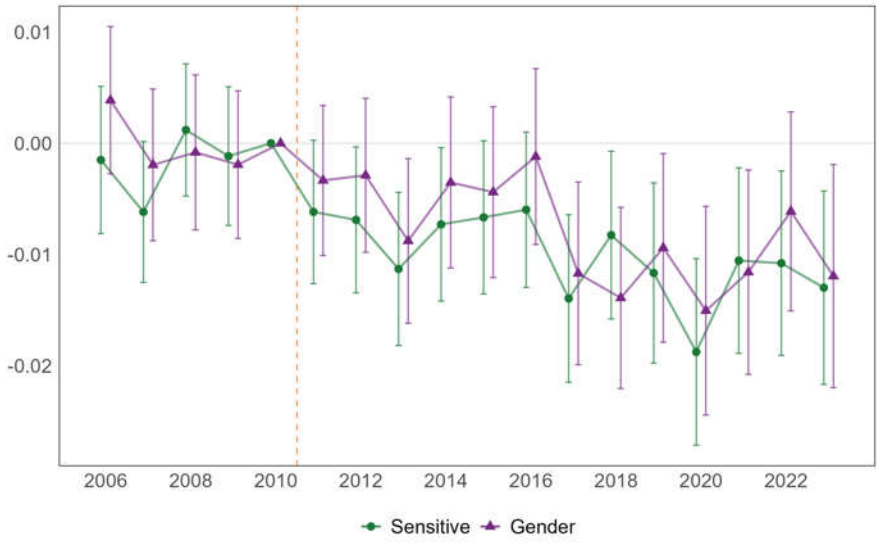
4. Between *Loyalty* and *Silence*: *Self-censorship*

We consider in this section that there might be an alternative strategy of academics, between *loyalty* and *silence*, which is likely to be *self-censorship*. Kováts and Rónay (2021) show qualitative evidence that academics may feel the need to self-censor in their research so as not to attract negative attention.

We focus in this section on researchers working on targeted topics, namely sensitive and gender-related topics, and analyse to what extent we can observe patterns indicating self-censorship among these researchers. While this phenomenon is by definition hard to measure, we attempt to do so by looking at the textual content of research papers, especially looking at how researchers in targeted topics are more likely to change research direction after 2010 and publish research less similar to their earlier work. In order to do so, we estimate the same specification as before, using the targeted-topic treatment defined in Section V.1.b: a researcher is treated if they have at least three publications in the targeted topics prior to 2005. The main outcome of interest is the cosine similarity between a researcher's current publications and their own research published in 2005 or earlier[41]. A positive effect on this measure would indicate that treated researchers "go deeper" into their existing research direction, or into the way they frame their research, relative to neutral researchers. Conversely, a negative effect would indicate that they deviate from their previous research more than control researchers do, which could indicate either self-censorship or alternatively alignment with the research agenda of ally researchers.

[41] For robustness, we build similar measures using the similarity with a researcher's own publications up to 2010, between 2000 and 2005 and between 2005 and 2010.

Figure 1.7 : Effect on Research Similarity with pre-2005 Publications - Gender and Sensitive Topics



Notes: The graph above reports Fixed-Effect OLS regression estimates for the difference in the similarity measures for researchers publishing in gender related topics or in sensitive topics, compared to other Hungarian researchers for the period 2006-2023. Authors are matched using their number of publications over the period 2000-2005, their main field of publication and their year of entry in academia.

Table 1.7 reports the aggregated treatment effects for researchers working on gender-related topics, both pooled over the post-2005 period and separately by government mandate. We identify different patterns. First, the similarity of researchers' Hungarian-language publications with their own pre-2005 research declines significantly after treatment as shown in the first two columns, with the magnitude of the effect growing across mandates, from −0.004 during the first mandate to −0.012 during the fourth. Second, as can be seen in the next two columns, the similarity of their English-language publications with their pre-2005 research remains stable, and if anything increases slightly: treated researchers do not seem to abandon their original research agenda in their international publications. Third, similarity with the research of government-allied researchers also decreases, again with a magnitude that grows over successive mandates (see columns (5) and (6)).

Taken together, these results indeed point to an "in-between" strategy consistent with self-censorship[42]. Targeted researchers neither converge towards the research of political allies nor fully maintain their previous research direction in domestically visible output. Instead, they appear to adjust the content of their Hungarian-language publications away from their earlier, politically sensitive work, while preserving their research agenda in English-language, internationally oriented publications. Figure 1.7 shows that the corresponding event-study estimates display no differential pre-trends, supporting a causal interpretation.

These findings hold when replicating the analysis using working on sensitive topics as treatment, as reported in Section 1.G.4. We obtain comparable results in magnitude for these researchers, whereas no such pattern emerges for topics that do not seem as politicised in the Hungarian context, such as research on climate change.

[42] An alternative explanation would be that journal editors became less willing to publish works on these topics. We detail in Appendix 1.E.5 why this mechanism is likely not the one dominating: at the journal level, the share of publications on sensitive and gender-related topics is highly correlated before and after 2010, and the quality of the Hungarian and international journals in which treated researchers publish does not decline.

In addition, we estimate a triple difference-in-differences specification comparing researchers publishing on targeted topics in Hungary with their counterparts in the other countries for which we collect CRIS data, namely Slovenia and Croatia. These additional results confirm that our findings are not driven by a general trend in research topics: researchers working on gender in Hungary become less likely to continue publishing on the topic, relative both to Hungarian researchers in other fields and to gender researchers in the comparison countries. Figure 1.G.1 shows the event-study estimates of the triple interaction for the probability of publishing work labelled as *gender* after 2010. Targeted researchers experience a sharp decline in this probability, of around 4 pp on average, with the effect growing in magnitude over time.

Table 1.7 : Researchers Working on Gender: Research Similarity

	sim. pre-2005 (HU)		sim. pre-2005 (EN)		sim. Ally	
	(1)	(2)	(3)	(4)	(5)	(6)
gender × post	-0.006***		0.001**		-0.005**	
	(0.002)		(0.0005)		(0.002)	
gender × post (1rst m)		-0.004**		0.0006		-0.0004
		(0.002)		(0.0004)		(0.002)
gender × post (2nd m)		-0.007***		0.001*		-0.007**
		(0.003)		(0.0005)		(0.003)
gender × post (3rd m)		-0.010***		0.0008		-0.013***
		(0.003)		(0.0006)		(0.004)
gender × post (4th m)		-0.012**		0.002**		-0.014**
		(0.005)		(0.0009)		(0.006)
Mean dep. var. (pre)	0.819	0.819	0.843	0.843	0.771	0.771
Observations	204,954	204,954	164,034	164,034	93,887	93,887
Author fixed effects	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓

Note: This table gathers the FE OLS regressions for gender-topic researchers. Columns (1)–(2) report estimates for the similarity of Hungarian-language publications with the researcher's own pre-2005 publications, columns (3)–(4) show the estimates for the similarity of English-language publications. Columns (5)–(6) for the similarity with the current research of government-allied researchers. Odd columns report the average post-2010 treatment effect and even columns report the decomposition of this effect by mandate. All specifications include author, cohort × year and field × year fixed effects. *** p<0.01, ** p<0.05, * p<0.1. *** p<0.01, ** p<0.05, * p<0.1.

VII — What Is the Global Effect of Illiberal Policies?

The comparison made in the previous sections identifies the differential effect of political alignment, but it nets out any effect common to all Hungarian researchers, including neutral ones. In this section, we estimate the illiberal penalty that all researchers may bear by comparing neutral Hungarian researchers with matched researchers in two neighbouring countries, Slovenia and Croatia. These estimates capture the effect of the post-2010 policy environment on Hungarian academia as a whole, rather than its differential impact across groups of researchers.

1. Comparing Individual Research Outcomes between Countries

In this section, treated researchers are neutral researchers based in Hungary, i.e. researchers whose collaboration index with both political opponents and political allies is zero under the network definition, and control researchers are researchers based in Slovenia or Croatia. We keep researchers who entered academia between 1990 and 2005. Each treated researcher is matched to control researchers following the procedure described in Section IV.1.b.[43] We thus estimate the following event-study specification:

$$Y_{it} = \sum_{\substack{k=2006 \\ k\neq 2010}}^{T} \beta_k \, T_i \times \mathbb{1}\{t = k\} + \alpha_i + \delta_t + \sum_{s=2006}^{T} \gamma_s \, X_i \times \mathbb{1}\{t = s\} + \varepsilon_{it}, \tag{1.3}$$

where Y_{it} is an outcome of researcher i in year t, T_i equals one for neutral Hungarian researchers, and α_i and δ_t are researcher and year fixed effects. X_i is a set of controls, namely field and cohort indicators, so that the terms $\gamma_s X_i \times \mathbb{1}\{t = s\}$ correspond to field × year and cohort × year fixed effects. The year 2011 is again the first treated year and 2010 is the reference year. Each β_k measures the difference between treated and control researchers in year k relative to 2010, and the coefficients for 2006–2009 test for parallel trends. To summarise the dynamics, we also replace the yearly interactions with interactions for each of Viktor Orbán's mandates.[44][45] Our estimates capture all changes that affected neutral Hungarian researchers relative to their neighbours after 2010, and all other Hungary-specific shocks are therefore part of the treatment.[46]

2. Research Productivity and Research Quality

Table 1.8 reports the estimates obtained with national repositories. After 2010, neutral Hungarian researchers publish more than matched researchers in Slovenia and Croatia (column 1). The effect is already significant during the first mandate (+21%) and grows over time, to +43%

[43] We describe global cross-country trajectories and why comparing Hungary with other Central European countries is valid in Appendix 1.B.4.

[44] So we consider from 2011 to 2014 for the first mandate, then 2015–2018, 2019–2022 and 2023. For OpenAlex outcomes, the window ends in 2022. Count outcomes are winsorised at the 99th percentile and estimated by Poisson pseudo-maximum likelihood, and binary outcomes are estimated by OLS. Standard errors are clustered at the country × field level.

[45] In this section, treatment is assigned at the country level: all neutral Hungarian researchers are treated at the same time and are exposed to the same national shocks, so clustering by researcher would understate the standard errors. Clustering by country would match the level of treatment, but with only three countries, one of which is treated, cluster-robust standard errors are unreliable. We therefore cluster standard errors at the country × field level, which allows errors to be correlated among researchers of the same discipline in the same country, who share funding panels, journals and evaluation rules, while keeping a sufficient number of clusters. Since field × year fixed effects already absorb shocks common to a discipline across countries, this level of clustering accounts for the remaining correlation within each country and discipline. It does not, however, account for shocks that affect all Hungarian disciplines at once. We assess this concern with placebo tests and random assignment of treatment in Appendix 1.I.2.

[46] Croatia's accession to the European Union in July 2013 may also have improved Croatian researchers' access to European funding and networks. Using two control countries limits this concern, and most of the divergence we observe seems to start before 2014. We also address this issue in Appendix 1.I.

during the second mandate, +37% during the third and +52% in 2023. Most of this increase comes from publications in the national language (column 3), which rise in every mandate, by 17% during the first and by 30 to 35% later. English-language publications increase much less, and the effect is only significant during the second mandate and in 2023 (column 2). In relative terms, the gap between the two languages is for instance moderate during the second mandate (+30% against +19%). In absolute terms, however, it is much larger, because national-language publications were more than four times as frequent before 2011: at pre-treatment means, the second-mandate effect corresponds to about 0.17 additional national-language publications per year, against 0.02 in English.[47] This pattern is consistent with the policies that rewarded publication in national outlets and in Hungarian (Appendix 1.A).

Higher output comes with lower quality (columns 6 to 8 and Figure 1.9a). The number of publications in top journals does not change during the first mandate, but declines sharply afterwards. During the second mandate, neutral Hungarian researchers publish 48% fewer papers in top-5% journals and 42% fewer in top-10% journals than control researchers. During the third mandate, these declines reach 75% and 59%. At the same time, publications in the 50% least-cited journals increase by 74% during the second mandate and 64% during the third. The decline in top publications therefore does not reflect lower English-language output, which does not fall in any period. Instead, the timing matches the introduction of the chancellor system in 2014 and the reorganisation of the research network after 2018. Effects on top journals are smaller and not significant in 2023. Since this period covers a single year and top publications are rare (about 0.02 to 0.03 per researcher and year before 2011), these estimates are imprecise. Lower funding could also explain lower-quality research. Hungary cut public spending after the 2008 financial crisis and during the European sovereign debt crisis, but higher education and research funding did not decline in aggregate, nor was it reallocated away from the social sciences (Figure 1.A.5). We lack researcher-level funding data and cannot estimate the role of funding directly.[48] Moreover, to the extent that funding cuts were part of the policy package, they are part of the effect we measure rather than a confounder. Overall, neutral Hungarian researchers became more productive after 2010, but mostly in Hungarian and at the expense of their international visibility: from 2015 onwards, they publish much less than comparable foreign researchers in the top journals of their field.

Figure 1.8 plots the effects obtained when estimating the model using total publications in the top 10% cited journals as the outcome on researchers in each field separately. Similarly to above, the effect does not appear to be dependent on funding. In particular, fields that depend strongly on funding based on the estimation of Bertolotti, Myers, and Tham (2025) do not appear more affected than the others. Researchers working in fields that are typically not politicised such as mathematics or agriculture are also harmed.

3. Decoupling of National and International Collaboration Networks

We now focus on the effect of illiberal policies on researchers' networks. Table 1.8 shows that, after 2010, neutral Hungarian researchers publish more frequently with other researchers from their own country than control researchers do.[49] The coefficients reported in this table are obtained with CRIS data, and therefore only capture the effect on the *national* network of Hungarian researchers. This holds both for the total number of co-authors and for the number of new collaborations (columns 4 and 5). There is no effect during the first mandate, but the effect is large from the second mandate onwards: neutral Hungarian researchers have 58% more co-authors and 53% more new collaborations than Slovenian or Croatian researchers. At the pre-treatment mean, this corresponds to about 0.26 additional co-authors per year, i.e., around

[47] Pre-treatment means are computed on the estimation sample of each column, which differ because PPML drops researchers whose outcome is always zero.

[48] See for instance Kováts and Rónay (2021) for a comparison of higher education funding.

[49] See Figure 1.9b for the corresponding event-study estimates.

Table 1.8 : Effect on Research Quantity, Quality and Network

	Quantity			Network		Quality		
	All pub. (1)	English pub. (2)	Native-lang. pub. (3)	Co-authors (4)	New collab. (5)	Top-5% (6)	Top-10% (7)	Bottom-75% (8)
Mandate 1 × Hungary	0.188***	-0.020	0.155***	0.035	-0.034	-0.229	-0.183	0.119
	(0.037)	(0.062)	(0.052)	(0.054)	(0.055)	(0.237)	(0.211)	(0.130)
Hungary × Mandate 2	0.354***	0.171**	0.260***	0.458***	0.426***	-0.653*	-0.541*	0.555***
	(0.049)	(0.072)	(0.069)	(0.062)	(0.074)	(0.342)	(0.298)	(0.129)
Hungary × Mandate 3	0.315***	0.053	0.274***	0.509***	0.524***	-1.40***	-0.897**	0.494***
	(0.076)	(0.100)	(0.086)	(0.106)	(0.128)	(0.517)	(0.403)	(0.153)
Hungary × Mandate 4	0.417***	0.257*	0.302***	0.789***	0.859***	-0.096	-0.288	0.767***
	(0.094)	(0.142)	(0.099)	(0.129)	(0.161)	(0.415)	(0.270)	(0.160)
Observations	74,122	51,606	66,596	64,517	62,872	8,951	13,778	37,595
Pseudo R^2	0.39213	0.38730	0.39710	0.42381	0.30825	0.38108	0.35380	0.30058
Pre-mean (treated)	0.568	0.122	0.565	0.444	0.318	0.019	0.029	0.048
Researcher fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-Field fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-Entry year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

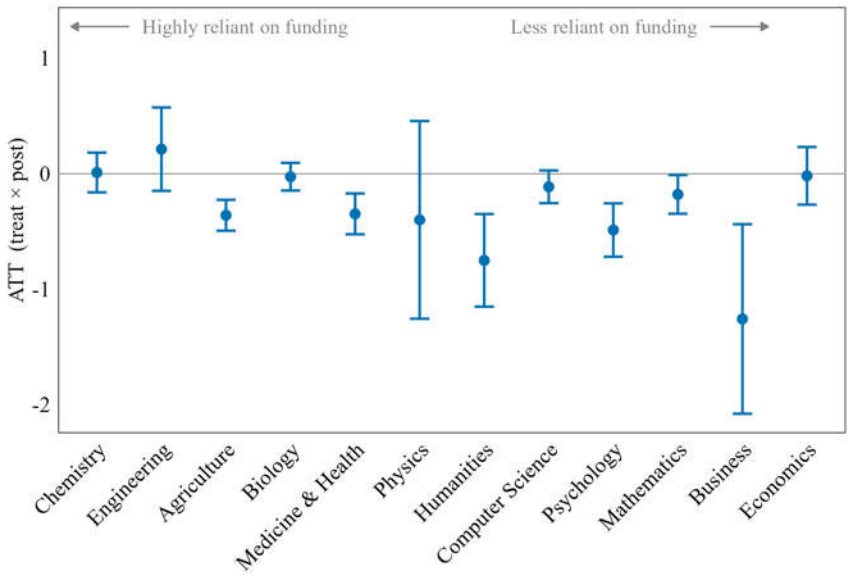
Note: Estimates of the mandate version of equation (1.3) by fixed-effects PPML on the cross-country sample in national repositories. Treated authors are neutral researchers based in Hungary and control authors are matched researchers based in Slovenia and Croatia. Top-5%, top-10% and bottom-75% refer to the citation rank of the journal. Pre-mean is the 2006–2010 mean of the outcome for treated researchers in the estimation sample. The number of observations varies across columns because PPML drops researchers whose outcome is always zero. Standard errors, clustered at the country × field level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

1 every four years. The effect grows further during the third mandate (+66% and +69%). Since the collaborations we capture here are by definition only collaborations with other "national" authors, this increase reflects a densification of the domestic network. Taken alone, this result is not evidence of isolation; it becomes so only when combined with the effect on international collaborations. We test this hypothesis using the probability of publishing with at least one co-author based abroad. International collaborations decline for neutral Hungarian researchers relative to control researchers (Table 1.9, columns 5 and 6). After 2010, the probability of publishing with at least one co-author based abroad falls by 5.9 percentage points. As for the national network and research quality, the effect is small and not significant during the first mandate, and appears from the second mandate onwards, with declines of 43% and 42% of the pre-treatment mean during the second and third mandates. This result is consistent with the effect we measure using national repositories. Together, they point to a decoupling between national and international collaborations: from 2015 onwards, the national networks of Hungarian researchers expand while their international networks contract.

4. Exit Through International Mobility and Leaving Academia

We then examine *exit* strategies of researchers compared to their control counterparts, using the cross-country sample in OpenAlex. We first find no effect on the yearly probability of publishing with an affiliation outside the home country (Table 1.9, columns (1) and (2)). However, the probability of publishing with an affiliation outside the home country for four consecutive years increases by 1.3 percentage points for neutral Hungarian researchers relative to control researchers (column 3). While this effect is small in absolute terms, it represents 26% of the pre-treatment mean of the variable (computed over 2006–2010). It is already significant during the first mandate and increases to about 30% of the pre-treatment mean during the second and third mandates (column 4). Overall, although the effect on emigration is moderate in absolute value, it is non-negligible relative to the baseline. This effect does not appear to be caused by an increase in double affiliations with a foreign institution: Table 1.H.2 shows that the effect is only significant for the probability to be both abroad and not in Hungary, while it is insignificant for the probability to be in both. Moreover, while we do not have the statistical power to fully assess where researchers emigrated, when splitting Europe in Western, Central, Northern or Eastern regions, Table 1.H.3 shows that the only positive effect, though

Figure 1.8 : Effect on Publications in top 10% cited journals by field



Notes: The graph above reports Fixed-Effect OLS regression estimates for the difference in the similarity measures for researchers publishing in gender related topics or in sensitive topics, compared to other Hungarian researchers for the period 2006-2023. Authors are matched using their number of publications over the period 2000-2005, their main field of publication and their year of entry in academia.

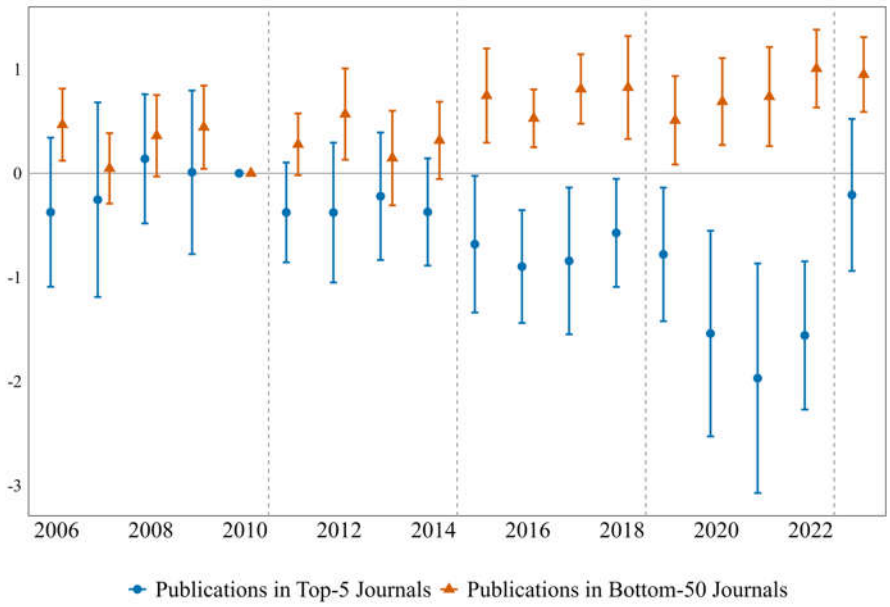
non-significant, is concentrated on Western Europe. This also means that the likelihood of spillovers to the control group from emigrating treated authors is low: there is no evidence that exit would lead them to move to Central Europe and collaborate with control researchers. Finally, we look at exits from academia. We measure them with an indicator equal to one in the year of a researcher's last observed publication.[50] Neutral Hungarian researchers are 0.77 percentage points more likely to exit each year after 2010 than control researchers (column 7). Relative to the pre-treatment exit rate of 2.7% among Hungarian researchers, this corresponds to a 29% increase. The disaggregated results show that this effect intensifies over time (column 8). It is small and not significant during the first mandate, reaches 0.66 percentage points (+25%) during the second, and 1.24 percentage points (+46%) during the third. This is again consistent with the increasing intensity of political pressure on academia.

5. Aggregate Losses from Political Pressure

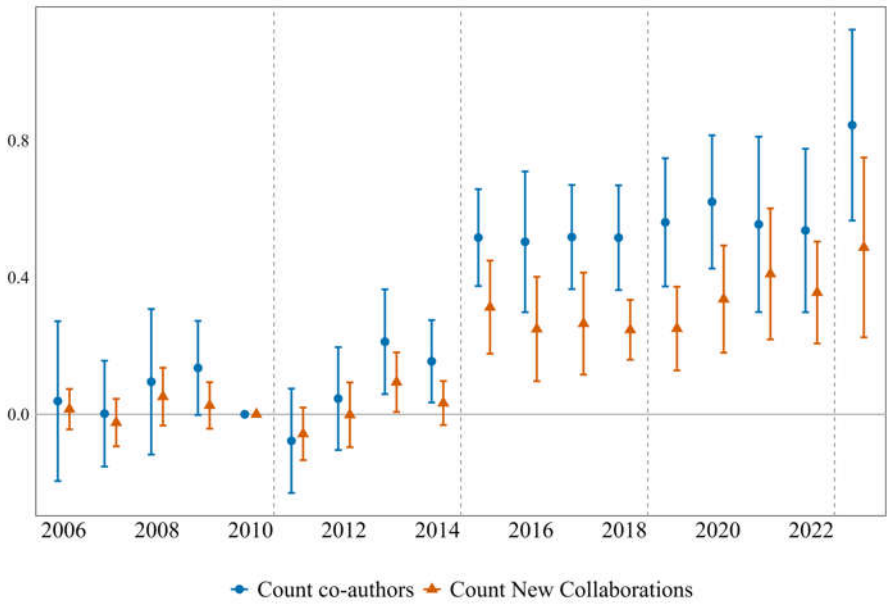
Beyond the estimation of micro-effects at the individual level, we assess the cumulative macro effect of illiberal policies on the aggregate output of research and on the stock of researchers over the entire period. Because our estimates are obtained on researchers who entered academia between 1990 and 2005, we focus on this population of researchers and apply our estimates to all treated researchers in it. Table 1.10 reports our calculations. For each outcome, the first four columns report the mandate coefficients estimated with equation (1.3) on the cross-country sample. Since most of them are obtained by PPML, column 6 converts them into an average percentage effect over the period 2011–2023. We apply this effect to the pre-treatment average of treated researchers (column 5), discarding any growth in academic productivity unrelated to the policy change, which gives the effect per person and per year (column 7). We then multiply this per-person effect by the number of researchers active in each year after the shock. A

[50] Since the exit from academia is measured using the last publication of one author, all researchers "mechanically exit" at the end of the sample. We thus use only data up to 2022 for this estimation.

Figure 1.9 : Effect on Publications by Journal Rank and on the National Co-Author Network



(a) Publications, by journal rank



(b) National co-author network

Notes: Event-study estimates of equation (1.3) by fixed-effects PPML on the cross-country sample in national repositories, 2006–2023, relative to 2010. The outcomes are (a) the number of publications by journal rank and (b) the total number of co-authors and the number of new co-authors registered in national repositories. Treated authors are neutral researchers based in Hungary; control authors are matched researchers based in Slovenia and Croatia. Outcomes are winsorised at the 99th percentile. Standard errors are clustered at the country × field level.

share of the researchers exits our database over the period, either because they no longer publish or because they are no longer in Hungary, and the effect no longer applies to them. Thus, we only count researchers who are active in each year. Column 8 reports the resulting cumulative effect for matched treated researchers, and column 9 for all treated researchers in the full sample. Column 9 assumes that the effect estimated on matched researchers also applies to unmatched treated researchers. Column 10 only keeps the mandates whose coefficient is significant at the 10% level, which gives a more conservative estimate. Finally, the last column expresses the

Table 1.9 : Effect on International Collaboration, Migration and Exit

	Affil. other country		Id. four years		Intl. Collaborations		Exit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
hungary × post	-0.0006		0.0134***		-0.0592***		0.0077***	
	(0.0035)		(0.0036)		(0.0221)		(0.0021)	
hungary × post (1rst m)		0.0040		0.0108***		-0.0156		0.0034
		(0.0036)		(0.0033)		(0.0178)		(0.0025)
hungary × post (2nd m)		-0.0014		0.0156***		-0.0810***		0.0066**
		(0.0044)		(0.0045)		(0.0275)		(0.0026)
hungary × post (3rd m)		-0.0047		0.0162***		-0.0791**		0.0124***
		(0.0060)		(0.0050)		(0.0317)		(0.0031)
hungary × post (4th m)								0.0098
								(0.0075)
Observations	203,938	203,938	203,938	203,938	203,938	203,938	114,354	114,354
R^2	0.467	0.467	0.698	0.698	0.533	0.533	—	—
Within R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mean Dep. Var. (Pre)	0.087	0.087	0.052	0.052	0.190	0.190	0.027	0.027
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: OLS estimates of equation (1.3), pooled over 2011–2024 (odd columns) and by mandate (even columns), on the sample of researchers who are matched between sources (national repositories and OpenAlex), 2006–2024. Treated authors are neutral researchers based in Hungary and control authors are matched researchers based in Slovenia and Croatia. Columns 1–2: indicator equal to one if the author publishes with an affiliation outside the home country in year t. Columns 3–4: indicator equal to one if the author publishes with an affiliation outside the home country for four consecutive years. Columns 5–6: indicator equal to one if the author publishes with at least one co-author based abroad. Columns 7–8: indicator equal to one in the year of the author's last observed publication. Regressions include author, field × year and cohort × year fixed effects. Standard errors, clustered at the country × field level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

total cumulative effect as a share of the counterfactual outcome over the period, i.e. observed publications minus the estimated effect. The rise in publications in the 50% least-cited journals translates into between 8,300 and 8,700 additional publications, or 17% of the counterfactual number of such publications. While the number of lost publications in the top 5% and top 10% journals is much lower, between 3,500 and 4,300 and between 4,200 and 5,600 respectively, they still account for a loss of 6% and 5% of all such publications. Given that these estimates are substantial, we believe that our paper documents a real shift in the composition, quality and international exposure of Hungarian research. We further estimate that illiberal policies caused between 1,700 and 1,800 additional exits from academia, which represent about 11% of the treated researchers active after 2010.

Table 1.10 : Aggregate Treatment Effects by Mandate

	Coefficient by mandate					Average over 2011–23		Cumulative effect			
	M1 2011–14	M2 2015–18	M3 2019–22	M4 2023	Pre-mean	Effect	Per-person/year	Matched	Total	Total (sig.)	Share
Top 5%	−0.23	−0.65*	−1.40***	−0.10	0.069	−41.4%**	−0.03**	−419***	−4,314**	−3,521***	−5.9%**
	(0.24)	(0.34)	(0.52)	(0.41)		(16.8)	(0.01)	(150)	(1,748)	(880)	(2.3)
Top 10%	−0.18	−0.54*	−0.90**	−0.29	0.102	−36.4%**	−0.04**	−522**	−5,569**	−4,249***	−4.9%**
	(0.21)	(0.30)	(0.40)	(0.27)		(16.5)	(0.02)	(227)	(2,525)	(1,424)	(2.1)
Bottom 50%	0.07	0.48***	0.48**	0.70***	0.116	50.1%**	0.06**	739**	8,735**	8,324***	17.1%**
	(0.15)	(0.15)	(0.20)	(0.21)		(22.6)	(0.03)	(358)	(3,937)	(3,181)	(9.0)
Exit[(a)]	0.00	0.01***	0.02***	0.01***	0.014	0.85*** pp	0.01***	200***	1,808***	1,719***	10.7%***
	(0.00)	(0.00)	(0.00)	(0.00)		(0.15)	(0.00)	(34)	(313)	(233)	(1.8)

Notes: Columns M1 to M4 report the coefficients of the interaction between the treatment (being in Hungary) and each mandate dummy, estimated by Poisson pseudo-maximum likelihood (PPML) on the matched sample. The estimation includes individual, year × field, and year × entry cohort fixed effects. Standard errors in parentheses are clustered at the country × field level. [(a)] Estimated with a linear probability model (LPM) and the effect is in percentage points, and cumulative effects in number of individuals. Cumulative effects are computed as $\sum_m (e^{\hat{\beta}_m} - 1) \times \bar{y}_{\text{pre}}^{\text{treated}} \times \sum_{t \in m} N_t$ for Poisson outcomes and $\sum_m \hat{\beta}_m \times \sum_{t \in m} N_t$ for LPM outcomes, where N_t is the number of active treated individuals in year t. *Total* uses all treated researchers with entry cohorts 1990–2005 in the full sample; *Total (sig.)* only keeps mandates whose coefficient is significant at the 10% level. *Share* expresses the *Total* cumulative effect as a percentage of the counterfactual number of publications of treated researchers over 2011–2023 (observed publications minus the estimated effect) for Poisson outcomes, and as a percentage of all treated researchers active after 2010 for LPM outcomes. *Per-person/year* is the total cumulative effect divided by the number of treated person-years over 2011–2023, and *Effect* expresses it relative to the pre-mean, i.e. $\sum_m w_m (e^{\hat{\beta}_m} - 1) \times 100$ with w_m the share of person-years in mandate m (for LPM: $\sum_m w_m \hat{\beta}_m \times 100$). Standard errors of aggregate effects are obtained by the delta method using the clustered variance-covariance matrix. Active individuals are researchers with at least one publication in the year (Poisson outcomes) or not yet retired (LPM outcomes). *** p<0.01, ** p<0.05, * p<0.1.

VIII — Conclusion

Our work shows how illiberal policies and political pressure can directly affect researchers and the production of ideas. We use the context of Hungary as a quasi-natural experiment to measure how such pressures affect the quantity and quality of research. We first show that researchers in Hungary are negatively affected by illiberalism when they are perceived as political opponents. These researchers experience a decline in both the quantity and the quality of their research, and are also more likely to leave Hungary. We document that these effect likely do not go through funding cuts, but rather come from the contraction of the perceived political opponents' network, suggesting that political pressure indeed makes collaboration more difficult for targeted individuals. Although silencing critics is often a central aim of illiberal regimes, targeted researchers have been more likely to publicly express their opposition to the regime, but also to self-censor in order to avoid the consequences of the political pressure.

Finally, we show that illiberalism imposes a penalty on all Hungarian researchers, including those who are not politically involved. Compared with similar researchers in other Central European countries, Hungarian researchers publish more, but mostly in Hungarian, and publish much less in the top journals of their field. We document an increasing isolation of the country's academia: Hungarian researchers collaborate more with their national colleagues, to the detriment of their international collaborations, and they are more likely to move abroad durably and to leave academia. Overall, illiberal policies shifted the composition, quality and international exposure of Hungarian research.

This work is particularly relevant at a time when rising political forces increasingly present the policies implemented in Hungary as a model. It contributes to understanding how not only the degradation of formal, legal institutions but also the political climate they create can foster, or deter, innovation and the production of new ideas. While our findings may not be fully generalizable to all contemporary political environments, the targeting of academics is often among the earliest signs of a shift toward illiberalism. Our results suggest that academia, as an organised system for producing innovation through individual researchers, remains sensitive to political pressure, even in regimes that occupy the middle ground between full autocracies and liberal democracies. The growing popularity, across democracies, of political parties that advocate similar regimes is therefore a serious concern for future innovation.

Chapitre 2
Characterising the Gains to Autonomy in the Public Sector: Evidence from French Academia

I — Introduction

As the cost of the European welfare states has risen, the demand for the public goods they provide has increased as well. There is no obvious solution to this dilemma. In most states, the common answer has been to attempt to transcend it by making the state more efficient. In practice, this process was implemented through bundles of reforms. Decentralisation into smaller units competing among themselves and delegation of budget under strict evaluation have been applied to different extents to all sectors of public service. Does this process really increase output in the public sector? In this paper, I leverage long-lasting heterogeneity in status in French academia to study an application of this framework to French universities in the late 2000s. The new salience of international competition, in particular through the Shanghai ranking, exposed inefficiencies in French academia, which was then reformed with the explicit goal of improving both its research output and its international standing. The French university system was fully centralised under the Ministry of Higher Education and Research; the reform decentralised it at the university level. Funding was allocated fully on the basis of predicted needs; a new channel for funds was opened to be allocated based on research outcomes and proposals, with both institutions and researchers having to compete to receive it. As more impactful research projects need more money, the new allocation may be more efficient. On the other hand, managing the university's assets and its workforce, allocating money that was directed prior to the reform, and applying for internal grants costs time that was free prior to the reform and allocated to other activities. I find that treated institutions have managed to improve their international exposure and attractiveness, increasing their stock of researchers by 15%, and their total output publications by 23%. However, individual researchers affiliated to universities prior to the reform lose on average 40% of their pre-reform publications per year after switching to the new system.

Identification of the effects of the reform relies on a historical feature of French academia. In France, universities are not the sole producers of publicly funded research: all of the other institutions, namely the *Grandes écoles* and the *Grands Organismes de Recherche* ecosystem, for instance INSERM[51], were left structurally untouched by the organisational reform only due to their different administrative status. These institutions employ a substantial share of French researchers, publish in the same journals and fields as treated universities, and were subject to the same broader trends in European research policy over the period: they were not excluded

[51] *Grandes écoles* are selective elite institutions historically focused on teaching, outside the public university system. *Grands Organismes de Recherche* are public institutions that directly depend on funding from the state independently of universities.

from the reforms due to their research trajectory.

The 2007 Law Relative to Liberties and Responsibilities of Universities (*Loi relative aux Libertés et Responsabilités des Universités*, hereafter LRU) was intended as a large, structural change to how academics conduct their work in the country. First, it granted to universities the responsibility to manage their workforce[52], which had been directly managed by the Ministry of Higher Education and Research, as well as their assets and their budget. The intention was to let them choose themselves which projects to focus on, allowing competition for resources within institutions[53], and allowing them to merge into larger universities[54]. Second, it changed the hiring process. Committees for the allocation of new positions became less rigid. Critically, it allowed universities to create non-tenured contracts, intended to attract young and international researchers for short periods of time, given that the labour market for university professors in France, as in much of Europe, is a public sector market with strict rules[55]. The legislator expected effects on two margins: better-performing universities could attract better researchers using their budget and new types of positions, and they could better incentivise researchers to work more efficiently. An opportunity was then provided to reap the benefits of good governance projects: Excellence Initiatives ("Initiatives d'Excellence", hereafter IDEX), introduced in 2010, created France-wide contests for funding for groups of universities and private institutions based on excellence for both research and teaching, using 8 billion euros from post-Great Recession loans. IDEX competitions were based on clusters of universities: therefore, competing depended heavily on neighbours, and many institutions with strong research output did not compete. The two reforms were implemented sequentially as complements and affected institutions that employ the largest share of higher education and research professionals in France. The reforms do not fully overlap: 55% of institutions treated by the LRU received an IDEX, compared to 11% of institutions that were not treated by the LRU. Institutions that managed to convince the IDEX jury early obtained funding in 2011 and 2012. A second wave of the contest, including lower levels of funding (I-SITE) was held in 2016.[56] Universities that secured funding in each period differ both from each other and from those that never did: therefore, I also separately analyse winners of each wave as well as those universities that never received funding from the competition.

In order to assess the effects of the reform on productivity in public research, I use bibliometric data from OpenAlex, over the years 2003 to 2025. It contains information related to publication outcomes, such as total number of publications and citations, which can also be linked to text-based quality metrics computed by Arts, Melluso, and Veugelers (2025), and about affiliation of a researcher. This allows me to track the productivity of 290,305 researchers linked each year to 1,739 institutions. By aggregating at the institution level, this dataset also lets me track the mobility effects of the reform, both globally and by seniority and early-career productivity, for 1,948 institution-domain pairs. Treatment status comes from official decrees

[52] I assume in this context that universities care about professors primarily as researchers, and that they regard their teaching ability second, as attractiveness of university careers and the potential brain drain towards other countries had been a concern of lawmakers since 2001 (Fréville, 2001).

[53] The LRU also reformed the system for university funding to make it nominally dependent on performance measures for 20% of its total. In practice, a French Senate report from 2019 finds that funding allocations changed very little, with most resources still originating from the annual public service grant (Adnot, 2019).

[54] Due to small sample size in mergers, only 16 during the 2007-2026 period, I cannot analyse the effect of this change.

[55] After the reform, a subset of the available positions is decided by the university in a discretionary manner, from a dedicated budget envelope. In particular, the law allows university presidents to hire workers with non-permanent contracts, explicitly in order to allow for more post-doc hires or temporary positions for foreign professors, but also to employ temporary workers for teaching purposes. Universities could not directly set researcher wages, which remained bound by national public-sector salary scales, but a newly authorised university foundation structure allowed them to complement those wages from privately raised funds.

[56] Two IDEX (Toulouse in 2016 and Lyon in 2020) and one I-SITE (Besançon in 2021) were cancelled after part of the money was already received. I do not include them in the analysis.

for autonomy and from ANR open data for IDEX, matched to affiliations. The reform is incorporated in a staggered manner: I estimate the effects following Callaway and Sant'Anna (2021), comparing treated researchers and institutions to comparable never-treated counterparts.

First, the reform appears to have achieved its intended purpose of allowing universities to attract more scientists. In all treated institutions, the number of active researchers rises by around 3.9 after 6 years, which is very large as it represents around 15% of the pre-period average. This is in part reflected in institution-level productivity: the number of publications rises steadily after treatment, by 11 publications on average per year over the period and 9 when only taking into account publications of researchers who arrived at the institution in the past five years. The increase in the number of foreign entrants is positive but insignificant, yet there is still evidence that they are more numerous as publications by this type of researchers increases on average by 5 per year over the period. Conversely, there is no significant effect on the average number of publications at the institution level. Average measures of quality also feature this rise in total output and lack of change or decrease in average output, including citations and number of publications in the top 5% cited papers. Institutions that received IDEX at either stage do not differ significantly in their effect from either the others or the pooled estimate, and the trends they follow are all similar. The absence of heterogeneity in effects is surprising, in particular in productivity and attractiveness to foreign researchers, as competition should favour the winners of the first IDEX as their project for governance was selected as promising. It suggests that the changes that durably affected French universities are the ones common to all in hiring rather than the possibility of leveraging autonomy to achieve excellence in hiring.

The other way to increase measured performance is to make all researchers more productive regardless of composition. The LRU also gave university management a variety of tools to raise it: access to private funding and increased wages for researchers through foundations[57] and excellence premia. However, the largest incentive came through IDEX, emulating the German Excellence Initiative programs of 2005. I find the total output of researchers affiliated to a treated institution in 2007, measured in publications, decreases by about 1.5 publications per year. Quality measures do not follow the same pattern: citations, publications in the 5% most cited journals and new phrase combinations weighted by reuse all drop during the first ten years after treatment before rising again to a level above pre-reform. Over the whole period, however, the negative effect is stronger both for citations and top 5% journal publications, at around a quarter of the pre-reform average, while it is offset by the second part of the curve for the text-based measure. Again, the time that the IDEX label was given by the international jury does not appear to matter: quality of the project does not appear to have an effect. Together, these results appear to show that while autonomy strongly increased the attractiveness of French universities to productive researchers, in particular from abroad, an administrative cost was borne by incumbent researchers even in the universities which had a high quality excellence project rewarded by an IDEX. While a share of the results, namely 31% of the effect for publications, is caused by an increased propensity to leave academia, the majority of the effect is driven by researchers who keep publishing. Researchers who were less research intensive prior to the reform, as measured by citations to 2000-2002 papers, are proportionally more affected than more research intensive counterparts. These patterns are consistent with a Holmstrom and Milgrom (1991) model of multitasking: by rewarding effort toward administrative tasks – in particular, in the context of IDEX, on grant applications, which are a partial signal of research activity – the reform diverts effort away from research itself.

This paper mainly draws on two fields of the economic literature. First, it contributes to the literature on the empirical effectiveness of organisational reforms in public services. A

[57] The foundation can either be a subdivision of the university itself used to gather funds from donors, or a private association directly managing them.

large theoretical literature establishes the optimality of competitive contracts for the provision of public utilities (Baron and Myerson, 1982; Laffont and Tirole, 1986; Laffont and Tirole, 1987; Pint, 1991), and empirical work has largely confirmed the gains from competition in liberalised settings, for water provision (Wolak, 1994), public transportation (Gagnepain and Ivaldi, 2002; Gagnepain, Ivaldi, and Martimort, 2013), and electricity (Fabrizio, Rose, and Wolfram, 2007). Evidence on more traditional public services is thinner and more mixed (Levin and Tadelis, 2010), though Cooper et al. (2011), Gaynor, Moreno-Serra, and Propper (2013) and Bloom, Propper, et al. (2015) show that increased competition in UK public healthcare raised productivity, improved managerial quality, and reduced mortality rates. Relative to this literature, I study a setting in which output cannot be contracted upon and where the wage-setting margin is largely closed and find results that contradict this positive effect at the individual public employee level.

Second, university organisation and organisational determinants of research quality have also been the subject of a very active empirical literature. Akcigit, Baslandze, and Stantcheva (2016), Murray et al. (2016) and Ganguli (2017) in particular show that researchers respond to tax incentives in their mobility, intellectual property in the direction of their research and funding changes in their productivity. This paper quantifies to what extent researchers also respond jointly in their mobility and productivity to changes in institutional rules. In an empirical paper closely related to this one, Aghion, Dewatripont, Hoxby, et al. (2010) show that in the US and Europe, more autonomous and competitive universities are better able to utilize their funds and generate on average more and better research than others. I show that the relationship between autonomy and research productivity does not appear to be clear-cut in the European context in which universities have less freedom to attract funding. I also contribute to the branch of the literature that focuses on the effect of receiving competitive funding. Azoulay, Graff Zivin, and Manso (2011) and Gush et al. (2018) both show that selective funding for top scientists at the individual level causes a large positive effect on their citations, while I find much more moderate effects. Conversely, Veugelers, Wang, and Stephan (2025) show that the ERC funding tends to select against researchers with more novel ideas and that applicants take on less novel ideas after being rejected, which is also the case for another similar program in Switzerland (Ayoubi, Pezzoni, and Visentin, 2021); my results corroborate these analyses by showing moderate effects of exposure to competitive funding in France on quality, but I also find that they increase text-based measures of novelty among recipients. Corsini and Pezzoni (2023) and Corsini (2022) also show that French excellence-based grants have moderate effects on researcher productivity, but do not measure the effect on novelty. Similarly, Möller, Schmidt, and Hornbostel (2016), Menter, Lehmann, and Klarl (2018), Buenstorf and Koenig (2020) and Cantner et al. (2026) all find limited and/or temporary results for the effect on academic productivity of the German Excellence Initiative. I complement these analyses by looking at the effect of overall governance including on losers of such contests, taking into account university heterogeneity and disentangling the effect of Excellence Initiative funding from other types of changes. Bertolotti, Myers, and Tham (2025) empirically estimate funding efficiency at the individual level in researchers, using survey data from the US; I look at the decision to switch funding systems at the macro level. Finally, Hall and Mairesse (2025) analyse Matthew effects caused by the CNRS model of funding, whereas I analyse the effects of switching models further away from it.

The paper is organised as follows. I first recontextualise the reform and its content in Section II. In Section III, I provide information about the data I use and sample selection. I describe the empirical strategy in Section III.3. I lay out results on mobility and productivity at the institution level in Section IV, and then at the researcher level in Section V. I discuss the external validity and welfare implications of these results in Section VI and conclude in Section VII.

II — Context and Content of the Reform

The French higher education and research institutions (hereafter HERI) are divided into three broad types. First, government agencies form a cornerstone of the research ecosystem in France, in particular the main national research agency, the National Centre for Scientific Research (CNRS)[58], but also other specialised institutions such as the Atomic Energy Commission (CEA), National Health and Medical Research Institute (INSERM), Institute for Research in Computer Science and Automation (INRIA) or the Institut Pasteur, in which HIV was discovered. The second type of HERI, the *Grandes Écoles* (GE), has historically been the training ground for French elites[59]. Starting from the 1960s, these institutions incorporated research as a potential outcome of their formerly professional training. While most of these schools, as well as the most reputable, are public (*École Normale Supérieure*, *École Polytechnique*...), some are fully private (*École des Hautes Études Commerciales*...) and some have a mixed structure (*Institut d'Études Politiques*...). Finally, universities are historically the first research institutions in France. However, the successive governments favoured the CNRS until the 1960s for national organisation of research, and the mass integration of students since the 1980s has weighed on their ability to perform high-level research. In law and medicine, universities have historically held a near-monopoly on training, but they have suffered from competition by Grandes Écoles in a wide array of fields from physics to literature. Della Malva, Lissoni, and Llerena (2013) provide a clear summary of the processes at play since the 1960s in the field of research in France.

These three statuses are defined at the level of the establishment (university, GE or government), but many research units have several of them. Since 1966, there have been laboratories combining CNRS funding and personnel with university status and researchers. Lawmakers codified them over time, naming them *Unité Mixte de Recherche* (UMR) in 1990. In the list of all active labs in France in July 2024, 1,551 out of 4625 are UMR, 1,046 of which are affiliated with CNRS. Among the 3,145 research units affiliated with universities, 1,413 have a joint affiliation with the CNRS, while only 951 are affiliated exclusively with a university. In the analysis, I consider for each treatment that a lab is treated if at least one of its parent institutions is treated.

Prior to 2007, the Ministry of Higher Education and Research managed all funding allocated to each university. At the same time, each component of the university had a director with strong power over the research and education decisions, which diminished the power of the president. In theory, the LRU reform addressed this by decentralising the management of resources at the level of the university and increasing the power of the university president in setting goals in education and research. I provide more information about the architecture in Appendix 2.A. In this paper, I analyse three different aspects of the reform, which are rolled out in a staggered way during the same period.

First, the 2007 LRU law transferred the rights to manage its workforce directly to the university, vastly expanding the amount of resources it managed. This right to manage the workforce is called "Responsabilités et Compétences Élargies" ("further responsibilities and authority", hereafter RCE). While the State still decides the number of the public servants in the university, all non-tenured jobs are managed by the university itself, on its own funds. The vast majority of the budget of universities is still directly decided by the Ministry, although the law encouraged the creation of private university foundations. Both of these changes were explicitly set in place in order to foster the employment of high-level foreign researchers. It also raised the

[58] The CNRS, a multidisciplinary research agency, was created in 1939 to foster excellence in research. It was home to the scientists responsible for many important scientific discoveries, such as the Bell-inequality tests that won Alain Aspect the Nobel Prize in 2022.

[59] Admission is conditional on a competitive national exam usually passed after two or three years of preparatory classes.

performance premium available to deserving researchers. In this paper, I group these changes induced by RCE under the label of "administrative autonomy". Between 2009 and 2015, the law granted RCE to 114 institutions, with the vast majority in the first four years. Figure 2.B.1 shows that the composition of treated institutions is broadly comparable across those first cohorts, but not after that. The 2013 cohort includes only overseas universities of Antilles and French Guiana, French Polynesia and Réunion Island, as well as some public Grandes Écoles with particular status. The 2014 cohort consists of only three small institutions with few researchers, and the 2015 RCE cohort consists almost entirely of teaching-focused public engineering schools, whose research profiles are sufficiently distinct from those of earlier cohorts to threaten the parallel trends assumption. They are therefore excluded from the main analysis.

The formula deciding the funds given to higher education and research institutions was also reformed by LRU, but institutional reports argue that the change was only nominal[60]. However, following the 2008 crisis, the government used the large post-crisis national loan program in favour of economic recovery to establish a system of competitive funding. The National Agency for Research (ANR) was founded in 2005, in an effort to provide project-based financing. In 2010, in the context of increasing university autonomy and results-oriented financing, the government instituted an investment plan for higher education and research called "Programme d'Investissement d'Avenir" (Investment for the Future Program, PIA). For higher education and research, the plan was administered by the ANR[61] and the funds were to be attributed via a competitive process on a project basis[62]. Some of the funding was to be delegated to research facilities directly (LABEX[63]), while some was to be attributed to teaching. The funds were to be taken from a project-based grant given to a "community of establishments" called IDEX. Again, the government's rationale was to promote synergy between different types of HERI, so that a given IDEX is typically received by a university and some other types of HERI[64]. 174 institutions received funding through 24 IDEX grants between 2011 and 2018 representing on average €77.2 million. Three IDEX grants were cancelled after a few years but some money was still distributed. The largest waves in terms of funding and number of recipients are the 2012 and 2016 waves.

IDEX was intended as a reward for positive use of new autonomy in universities: 55% of institutions treated for RCE were also recipients of an IDEX, compared to 11% of institutions that were not treated for RCE. The timing of RCE adoption was determined by voluntary application and administrative processing by the Ministry of Higher Education and Research, while IDEX eligibility required forming a multi-institution consortium and submitting a competitive proposal to a different authority. Those two selection processes are related but distinct and do not perfectly overlap. This difference in treatment decisions explains why 20% of doubly-treated institutions received their first IDEX funding before autonomy. In this paper, I consider the judgement of this authority to be indicative of the extent to which the autonomy project relied on focusing on research of excellence. I summarise the full timeline of treatment adoption in Figure 2.1.

[60] See Appendix 2.A for details.

[61] Convention du 23 septembre 2010 entre l'État et l'ANR relative au programme d'investissements d'avenir (action « initiatives d'excellence »), LOI n° 2010-237 du 9 mars 2010 de finances rectificative pour 2010.

[62] The IDEX winners received an endowment of which they could only use the interest until their status was confirmed.

[63] Investissements d'avenir, convention État–ANR Action : « laboratoires d'excellence »).

[64] For instance, the Bordeaux IDEX contains the two merged universities in Bordeaux but also Bordeaux's Institute for Political Science Studies (IEP Bordeaux), Bordeaux INP, and the National School for Agronomic Science in Bordeaux.

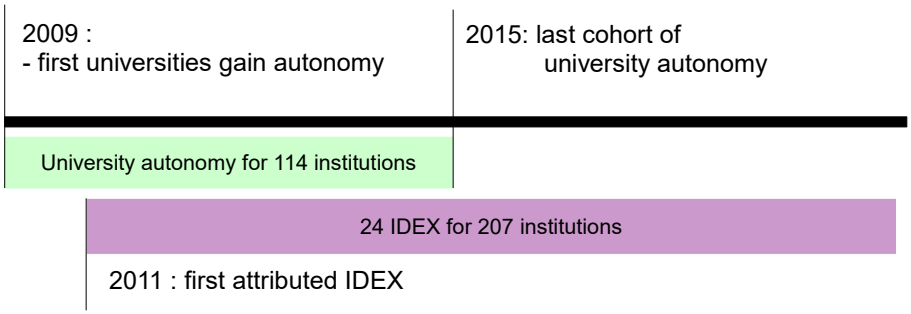


Figure 2.1 : Timeline of the Reform up to 2020

III — Data and Methodology

1. Data Collection

The dataset I use to measure both the mobility and productivity of researchers is OpenAlex, an open-access bibliometric database created in 2022 by a non-profit foundation using data from Microsoft Academic Graph. It is built at the level of papers retrieved from primary sources (mainly journals and publishers) and contains data about 250 million academic works and 90 million researchers (Priem, Piwowar, and Orr, 2022). For each publication, these data typically include a journal of publication, the university and country of affiliation of the author, and university-level and author-level characteristics. Compared to commercial alternatives such as Scopus or the Web of Science, OpenAlex provides the broadest coverage across journals, articles and citations (Hug, Ochsner, and Brändle, 2017; Visser, van Eck, and Waltman, 2021; Culbert et al., 2025). I leverage this information to build a dataset of French HERI, which I then use to identify authors affiliated with those institutions over the sample period, which spans 2003 to 2025.[65]

I construct treatment indicators from three administrative sources. I identify the year in which each institution received RCE from the annual implementation decrees published in the *Journal Officiel*[66] between 2009 and 2015. I derive which wave of IDEX the university took part in from open data provided by the ANR. I complement these with open-access data from the Ministry of Higher Education and Research to classify institutions by type: universities, *Grandes Écoles*, CNRS units, and other research organisations.

2. Selection of Samples

I build two samples covering the same French HERI. The first (hereafter *institution-level panel*) is constructed at the institution-field-year level in order to describe the effect of the reform on the evolution of the stock of researchers in labs and its composition. The second (hereafter *author-level panel*) is built at the author-institution-field-year level to describe productivity effects at the individual level.

Institution-level panel: The main difficulty in deriving an affiliation in bibliometric data is that an author is observed only when she publishes, but is also present at the institution in

[65] Results are robust to limiting the end period to 2020, the date of discontinuation of Microsoft Academic Graph (Alperin et al., 2024), after which data collection processes changed.

[66] The official repository of French government publications.

years during which she has no publication. For this sample, I avoid the issue by computing each year the number of active researchers, that is the number of researchers who publish each year. A second challenge is assigning fields of research to institutions, when some are multidisciplinary. Instead of seeing a given university in all fields as a single institution, I treat each institution's researchers in a given OpenAlex domain[67] as a unit. I define it as the main domain of publication of an author in OpenAlex. In other words, in my empirical analysis, "institution" means OpenAlex institution profile interacted with main domain of publication of an author. I keep French institutions with over 5 active researchers on average in the years 2000 to 2002, and with fewer than the 99th percentile of authors as they cannot be compared to other institutions. I then remove from the sample companies and hospitals, in which I consider the production function to be too different from more research-focused institutions[68]. I then split the stock of researchers based on observables. This allows me to compute the following variables to estimate the mobility margin of the reform: total number of researchers, total number of researchers by category of former affiliation (based on type of affiliation, or whether the affiliation was abroad or not), total number of researchers whose first publication was abroad, total number of researchers by gender, and by seniority. I also compute the same variable for all researchers who are "movers", meaning who were not present in the institution until this period, and "stayers", who were already active in the institution prior to this period. For all of these variables, I compute both the raw number of researchers that have an affiliation to the institution, and a weighted version in which each researcher with n affiliations contributes $1/n$ to the stock of researchers of each affiliation. This version is the one that is presented throughout the paper but results are robust to using raw counts. I exclude institutions that are not treated by autonomy but received an IDEX and those that were treated by autonomy after 2012, due to their incomparability with the rest as explained in Section II. The final sample contains 1,948 institution-domain units, among which 560 are never treated, 276 are treated in 2009, 564 in 2010, 303 in 2011 and 245 in 2012. The disproportion between the number of treated and control units is explained by the fact that many treated labs are also part of CNRS, but are mixed units, and that universities and CNRS are the largest employers of researchers in the public sector. I then measure research quantity with the total number of publications per year. I measure quality with the total number of citations, the number of publications in the top 20%, 10% and 5% of journals by field-year citation rank[69], and the average and total number of new phrases or words in publications using measures computed by Arts, Melluso, and Veugelers (2025)[70]. All variables are winsorised at the 1% level. I also compute the average value for all of these outcomes at the institution level.

Table 2.B.1 provides the average and median values for the most important outcomes for institutions during the pre-2007 period; Table 2.B.2 shows their distribution in the OpenAlex main domains, and Table 2.B.3 shows their split by type of institution in the database, share of CNRS lab, and whether they are in Paris. Institutions that were treated for administrative autonomy appear comparable across all those dimensions. While at first glance there appear to be strong differences between the control and the treated, Figure 2.B.2 and Figure 2.B.3 show that while control institutions are typically smaller, they follow parallel trends to the treated prior to 2007 without any controls, both in their increase in their number of researchers and their total production. Overall, the gap between the two appears to increase in all these metrics, suggesting that the effect of the reform would be positive.

[67] OpenAlex domains are broad fields of science: Life Sciences, Social Sciences, Physical Sciences and Health Sciences.

[68] I provide more information about the selection process in Appendix 2.B.

[69] I compute the citation and document count in OpenAlex for each journal that has over one publication per year, then I use this count to compute the journal rank in its field each year, similarly to Scopus's Citescore (James et al., 2019).

[70] The measure computes for each paper the number of combinations of phrases or words that have never appeared before in OpenAlex, weighted by their reuse.

Author-level panel: To build the second sample, I create a dataset at the publication-author-affiliation level in order to build a panel at the author-affiliation-year level. I now consider that an author is affiliated to an institution until she is observed to have been affiliated elsewhere[71]. This allows me to then aggregate at the author-year level. The variables used to determine research quantity and quality are the same as above; all variables are winsorised at the 1% level. I first keep researchers who had at least two years of activity and two publications in a French institution between the years 2003 and 2025, and whose first and last publications in OpenAlex are separated by at least two years. I then keep those who first published no sooner than 1965 and no later than 2003, in order to have a balanced panel. I then keep researchers who were active between 2000 and 2002, and those whose main field (including those who publish at the intersection of several fields) contains over 0.1% of the sample researchers. For the same reasons as above, I also exclude researchers who publish with an affiliation to a company or a hospital, or an institution in which the IDEX was cancelled. The final dataset is at the author-year level and comprises 69,143 researchers. Treatment status at the author level is assigned on the basis of treatment status of their affiliation in 2007 and can therefore be understood as intent-to-treat.

Table 2.B.4 reports descriptive statistics for the main outcomes analysed in this paper at the individual level. Compared to the treated, the control group clearly publishes less; however, they are comparable in novelty. Their distribution in domains, shown in Table 2.B.5, also appears to be comparable aside from health sciences, which is due to the removal of university hospitals from the sample.

3. Empirical Strategy

The reform has a staggered roll-out: autonomy does not affect institutions at the same time. In order to take this into account, I estimate the Average Treatment Effect on the Treated (ATT) for each treatment cohort and each period using the doubly-robust estimator of Callaway and Sant'Anna (2021):

$$ATT(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)\,C_t}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)\,C_t}{1-p_g(X)}\right]}\right)\Big(Y_t - Y_{g-1} - m_{g,t}(X)\Big)\right], \tag{2.1}$$

In which Y is the outcome, g is the treatment cohort and G_g takes value 1 for treated units and 0 otherwise, t indexes the calendar period, C takes value 1 for all control units and 0 otherwise, $p_g(X)$ is the propensity score estimated from a logit regression of the probability of belonging to cohort g, and $m_{g,t}(X) = \mathbb{E}\left[Y_t - Y_{g-1} \mid X,\ C_t = 1\right]$.

At the institution level, X includes the scientific domain of the institution-domain pair and an indicator for whether the institution is affiliated to the CNRS. Standard errors are clustered at the institution level (i.e., the level at which treatment is assigned). At the author level, controls include an indicator for activity of the author in each field, five-year bins of first year of publication, and the quartiles of publications and citations during the period 2000-2002. Coefficients are then re-aggregated following Callaway and Sant'Anna (2021) to generate the event studies and period-wide ATT presented in the paper. Standard errors are clustered at the 2007 affiliation level.

[71] If an author is first seen in institution A in 2004 and does not publish until 2007, in which she publishes in institution B, she is considered to be affiliated to A in each of these years.

In both cases, the control group includes both the units that have not yet been treated and the never-treated units. Given that universities had some ability to choose when they were treated, there could be selection on cohort of adoption; on the other hand, treated units are more alike compared to control units. Results hold when only comparing with never-treated units.

IV — How Did the Reform Affect Institutions?

This section describes the results of the empirical analysis of the effect of the reforms on French academia. I first show that autonomy has increased the attractiveness of universities to other researchers, to researchers both in France and abroad. I then show that these researchers increase the productivity of these institutions on average, but that this does not translate into higher productivity per researcher.

1. Autonomy Increased the Stock of Researchers

I first estimate the effect of administrative autonomy using Equation 2.1, with, as outcome, the number of researchers active each year at the institution in each of the four OpenAlex domains.

Figure 2.2 presents the average effect on the treated of administrative autonomy on the total number of researchers affiliated to treated institutions, weighted by their number of affiliations[72] by year relative to treatment. The stock of active researchers increases, and the effect becomes significant after 5 years. Over the whole period, the ATT represents about 15% of the pre-period average, and reaches a value of about 4 after 10 years, closer to 18%. The effect therefore sets in relatively fast and is large: relative to control and not-yet-treated institutions, treated institutions manage to significantly increase their hiring.

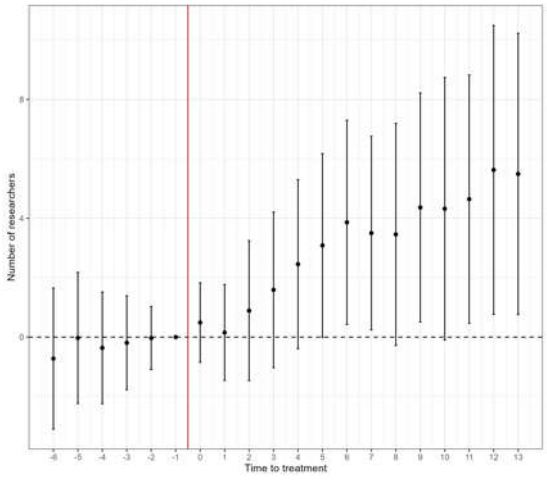


Figure 2.2 : Effect of Administrative Autonomy on Number of Researchers

Notes: This graph reports estimates of the effect on the total number of researchers per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

The legislator did not only intend to allow universities to attract a higher number of researchers, but also to make French academia more attractive to international researchers.

[72] A researcher with n affiliations counts with weight $1/n$ in this variable. Results are robust to using raw counts, as shown by Table 2.C.1.

While the data does not allow me to know the nationality of a researcher, I proxy it with the country of her first publication. Figure 2.3 shows that the reform also allowed university-affiliated institutions to increase their number of researchers whose first publication was abroad by about 1 after six years, in a way that appears permanent. While the effect appears small, there were only a little less than five such researchers in treated institutions prior to the reform, putting the effect in the same range as the global effect on attractiveness. The reform indeed appears to have increased openness to academics who started their career abroad, but not proportionately more than to academics who started their career in France. Results are robust to considering researchers whose previous affiliation was outside of France, as shown by Figure 2.C.1.

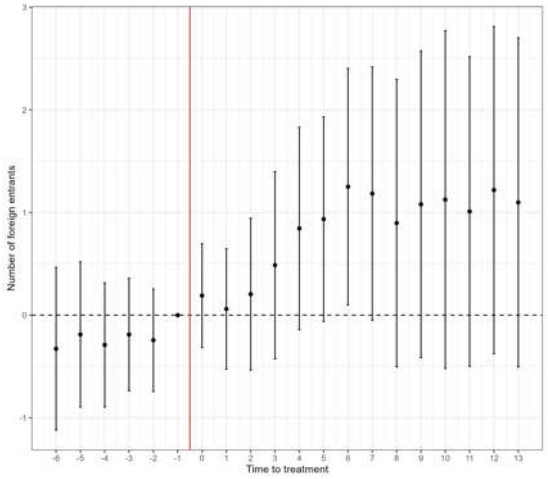


Figure 2.3 : Effect of Administrative Autonomy on Number of Researchers who First Published Abroad

Notes: This graph reports estimates of the effect on the total number of researchers whose first publication was published outside of France per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

Where the heterogeneity appears is in seniority of these hires. When splitting researchers by whether they are within five, five to fifteen or over fifteen years of their first publications, effects differ. Let us call those categories "junior", "mid-career" and "senior" researchers. The effect on the total number of junior researchers is insignificant, although the first panel of Figure 2.C.2 shows an insignificant hump-shaped curve, rising to one after five years and returning to zero after ten. Likewise, the effect on mid-career researchers is insignificant, with estimates closer to zero. Conversely, the number of senior researchers active at these labs rises steadily over time. Another indication of this compositional change is given by Figure 2.C.3, which shows that the number of researchers retiring each year increases as well. There is also no significant effect on the number of women active in treated institutions.

Table 2.1 summarises the ATT for these outcomes over the entire period. The effect on institution-level headcounts appears unambiguously positive, and the aim of attracting more foreign-based researchers appears to be fulfilled. However, this does not necessarily mean that those newcomers actually meaningfully produce research: they could be taking on some of the new administrative responsibilities accruing to university-affiliated institutions. Results are robust to truncating the panel to 2019, as shown by Table 2.C.2.

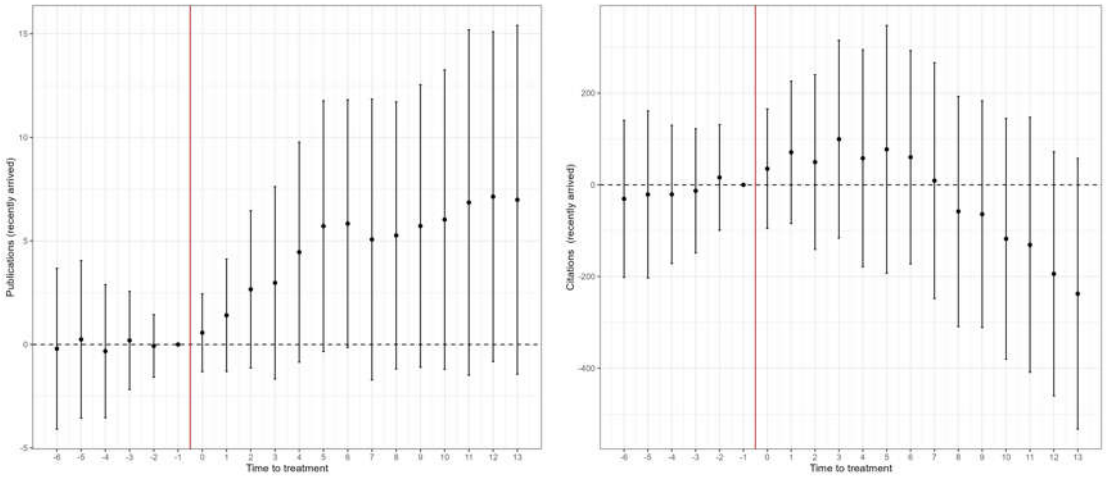
Figure 2.4a shows that this is not the case and that these newcomers contribute meaningfully to the total production of the institution. Publications by researchers affiliated to the lab in the past 5 years rise on average by 5.1 each year. While it appears to stabilize during the later years, it does not decrease, indicating that newcomers at each stage contribute to the

	Number of researchers (1)	*Number of foreign entrants* (2)	*Number of junior researchers* (3)	*Number of mid-career researchers* (4)	*Number of senior researchers* (5)	*Number of retiring researchers* (6)	*Share of women* (7)
All	3.532***	0.886**	0.140	0.179	1.967***	1.083***	-0.005
Std. error	(1.022)	(0.356)	(0.378)	(0.252)	(0.321)	(0.317)	(0.006)
Pre-treatment mean	22.290	4.746	9.095	4.507	4.829	2.433	0.236
N (units)	1,948	1,948	1,948	1,948	1,948	1,948	1,664
Controls							
Domain	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.1 : Effect of the Reform on Number of Researchers

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on the number of researchers in treated institutions per domain and per category: all (1), researchers who first published abroad (2), in the past 5 years (3), between 5 to 15 years prior to each year (4), and above 15 years (5), the number of researchers whose last publication was recorded this year (6) and the share of women (7). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

institution's research portfolio. However, the number of citations does not change significantly, as shown by Figure 2.4b. Similarly, Figure 2.C.4a shows that publications by foreign entrants rise significantly too, by 6 publications per year after 10 years on average, while their citations increase until about seven years after autonomy before returning to normal (Figure 2.C.4b). The effect over the whole period is still significant. Table 2.2 reports the ATT over year for these outcomes over the entire period. Publications by newcomers rise by 15%, while publications by researchers who first published abroad rise by almost 40%; their citations rise by about 32%. There is therefore evidence that the affected institutions benefitted from a positive size effect in their total production.



(a) Publications (b) Citations

Figure 2.4 : Effect of Administrative Autonomy on Total Production of Researchers with Less than 5 Years of Seniority at the Institution

Notes: This graph reports estimates of the effect on the total publications (left) and citations (right) at this institution for researchers who entered the institution in the past five years per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

Those affiliations also appear to be lasting: the effect on the number of new researchers or "movers", as shown by Figure 2.C.5a, is positive but not significant, while Figure 2.C.5b shows that the effect on researchers who were affiliated to the institution before or "stayers" follows the same trend as the total number. This also appears to be the case for researchers who first published outside of France. Table 2.C.3 presents the corresponding ATTs.

	Publications (recently arrived) (1)	*Citations (recently arrived)* (2)	*Publications by foreign entrants* (3)	*Citations by foreign entrants* (4)
All	5.098***	-58.933	4.375***	186.843**
Std. error	(1.906)	(63.906)	(0.916)	(85.640)
Pre-treatment mean	33.182	1310.186	11.435	579.895
N (units)	1,948	1,948	1,948	1,948
Controls				
Domain	Yes	Yes	Yes	Yes
Institution Type	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes

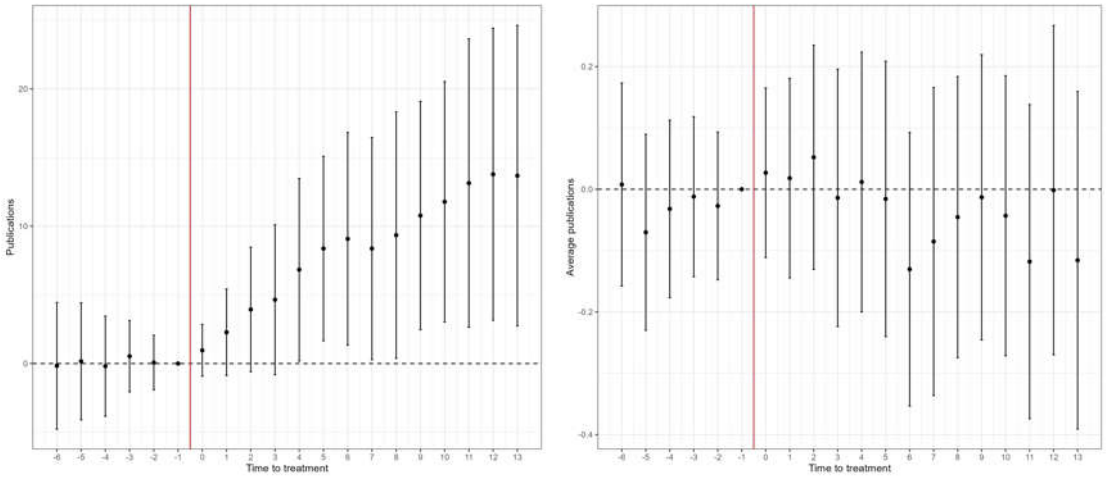
Table 2.2 : Effect of the Reform on Productivity of Newly Affiliated Researchers

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy at the institution-domain level on the number of researchers whose last affiliation prior to the current one was abroad, weighted by number of affiliations (1) and unweighted (2), and their number of publications (3). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

2. Autonomy Increased Total but not Average Productivity

The affected institutions appear to have increased their research productivity using newly affiliated researchers; in this section, I show that this effect cannot be extended to their entire stock of researchers.

First, the total number of publications in treated labs rises. Figure 2.5a shows that at the end of the period, a treated lab produces on average almost 14 more publications than before treatment, compared to a pre-2007 average of 38. Yet this effect disappears as soon as one considers the average number of publications, defined as the total number of publications divided by the affiliations-weighted number of researchers. Figure 2.5b presents the results of this estimation. The effect of the reform on this variable is not significant, and the point estimate is even negative. There are more people in each institution, but it does not appear that there are more publications per researcher.



(a) Total publications (b) Average Publications

Figure 2.5 : Effect of Administrative Autonomy on Total Production of Researchers at Treated Institutions

Notes: This graph reports estimates of the effect on the total publications (left) and average publications (right) of affiliated researchers for a treated institution per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

Figure 2.C.6a displays the estimates of the effect of receiving autonomy per year after treatment on the total number of citations received by the lab until 2025 for papers published in this year. It appears that this number is increasing, then decreasing; the effect is however insignificant. Because the number of researchers itself keeps rising, the point estimates for average citations in Figure 2.C.6b tend to be below zero, albeit non-significantly. For the number of new phrase combinations weighted by their reuse in OpenAlex, a text-based metric of Arts, Melluso, and Veugelers (2025), the pattern is similar: this number drops after 5 periods, this time significantly so when considering all these estimates together, but the average does not drop (see Figure 2.C.8).

I summarise the ATT for institution-level productivity in Table 2.3. The effect on publications reaches about 23% of the pre-reform average per year, while the non-significant negative effect on citations is not close to 0, at around 10% of the pre-reform average. While the effect over the entire post-treatment period on publications in the top 5% cited journals is not significant, Figure 2.C.7 shows that the effect is significant until year 8 after treatment, which coincides with the change in OpenAlex data collection practices. The effect until then reaches about 25% of the pre-reform average per year and is therefore sizeable. Yet again, there is no effect on average productivity using this metric.

	Publications (1)	*Average publications* (2)	*Citations* (3)	*Average citations* (4)	*Top 5% journal publications* (5)	*Avg. Publications in top 5% cited journals* (6)	*Avg #new phrase combinations* (7)	*Avg. New phrases* (8)
All	9.007***	-0.052	-13.975	-3.372	0.679	-0.045	-134.650	-1.148
Std. error	(2.463)	(0.063)	(76.338)	(3.505)	(0.875)	(0.039)	(87.963)	(4.532)
Pre-treatment mean	38.432	2.031	1490.818	81.696	17.464	0.915	2066.923	115.012
N (units)	1,948	1,664	1,948	1,664	1,948	1,664	1,948	1,664
Controls								
Domain	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.3 : Effect of the Reform on Productivity at the Institution Level

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy for treated institutions on total publications (1) and total citations (2) by researchers who became affiliated with the lab in the past 5 years, and total publications (3) and total citations (4) by researchers affiliated with the lab who first published abroad. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

The increase in institution-level productivity thus appears to be entirely driven by the mobility effect of the reform: affected institutions manage to attract more researchers, both from France and from abroad, but average productivity within these institutions does not rise. This is surprising, as one would expect that the addition of new productive researchers, in particular from foreign institutions, would have positive spillovers on their place of work as a whole; this suggests that either they are too small to be measured, or they are counteracted by another cause that decreases average productivity for other researchers.

3. Heterogeneity by IDEX Status and Field

In this section, I explore the degree to which the reform heterogeneously affected different institutions.

Let us first focus on a particular dimension of heterogeneity: the quality of the project of university management for the institution and its surrounding cluster. In order to assess it, I rely on whether the university obtained an IDEX in the first wave of funding, meaning that an international jury of researchers believed that the project was high-quality, in the second wave, meaning that there was a project to compete for resources that was eventually endorsed by ANR, or no IDEX funding, in which case the university either unsuccessfully applied, potentially at an unobservable cost, or did not apply, and we expect the university to be less focused on research and less competitive. Surprisingly, Figure 2.6 shows not only that the average effect on the total number of active researchers appears to be the same for all of these three types, but also that the

trends that this variable follows in all of them appear to be the same[73]. A consequence of a project pushing toward excellence could be reallocation: in other words, different laboratories belonging to first-wave IDEX winner universities could be affected differentially, leading the effect to be overall not different from institutions that did not qualify when considering them all together, with differences averaging out. This effect appears to be at play when considering the number of researchers. Figure 2.C.12 shows the effect on the institutions that received an IDEX in the first wave that lie respectively below and above the median number of publications in the selection period. While this is not significant, the effect appears to be higher for the institutions that lie above the median rather than below. However, this is not the case for production outcomes. In particular, Figure 2.C.11 shows that there is no significant difference in output. The absence of heterogeneity in productivity based on the quality of the project presented to the IDEX jury is puzzling, but it does not rule out heterogeneous effects of quality of project management at the researcher level.

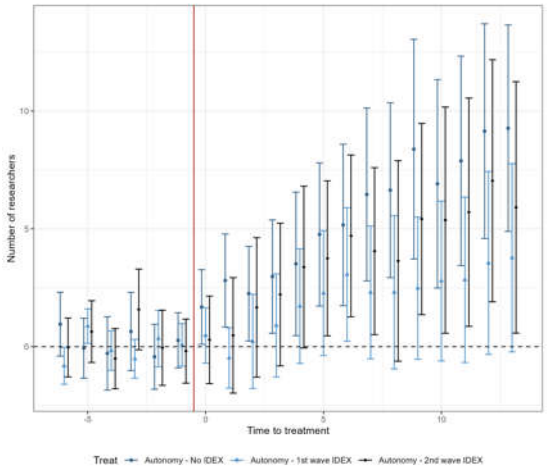


Figure 2.6 : Effect of Administrative Autonomy on Number of Researchers by IDEX Wave of the Institution

Notes: This graph reports estimates of the effect on the total number of researchers year after treatment for administrative autonomy. Estimates are obtained following Callaway and Sant'Anna (2021) in separate regressions for institutions that obtained an IDEX in the first wave (black), in the second wave (light blue) or never (dark blue). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

However, domains of study appear to be affected differentially: Table 2.C.4 shows that while the differences are not statistically significant due to large standard errors, the point estimates vary strongly in intensity for number of researchers, from about half of the pre-2007 average for social sciences to a fourth for life sciences. The effect on the number of foreign entrants is not significant neither for health sciences nor for social sciences. It is much stronger for physical sciences, at almost half of the pre-2007 average; however, in terms of total publications, Table 2.C.5 shows that it is the only domain for which publications do not significantly increase. The point estimate for average publications is negative for three out of four fields and is significant at the 5% level for social sciences. Conversely, institutions in life sciences are the only ones for which average citations appear to benefit from the reform; it appears that life sciences are the field that reaped the most from the reform across all outcomes, while the effect appears more marginal for health sciences. Interestingly, institutions in social sciences still mostly benefit from the size effect, but without any effect on attractiveness to foreign scientists, and with a

[73] The same closeness between those estimates from separate regressions - albeit using the same control group - can be seen in Figure 2.C.9 for total publications and total citations. The only outcome for which the gradient appears to matter, although still not significantly, as shown in Figure 2.C.10, is the number of new phrase combinations: higher-quality projects appear to be more negatively affected, which is in line with the results of Veugelers, Wang, and Stephan (2025) showing that competition for grants discourages novel research.

significant decrease in average production and both aggregate and average novelty. Surprisingly, while the effect on the number of researchers whose last affiliation was at a private company is not significant when estimated on the entire sample, it is positive and weakly significant for both social sciences and health sciences, and significant at the 5% level for life sciences. Finally, the share of women appears unchanged in all fields, with non-significant estimates close to 0.

V — How Did the Reform Affect Researcher Productivity?

I now assess the effects of the reform on individual productivity, by estimating Equation 2.1 on the author-level panel described in Section III.2. The panel is balanced and authors are considered treated when they were affiliated to a treated institution in 2007, so that the effect must be interpreted as an intent-to-treat, and as a local effect on researchers who were already active and present in affected institutions when the reform occurred.

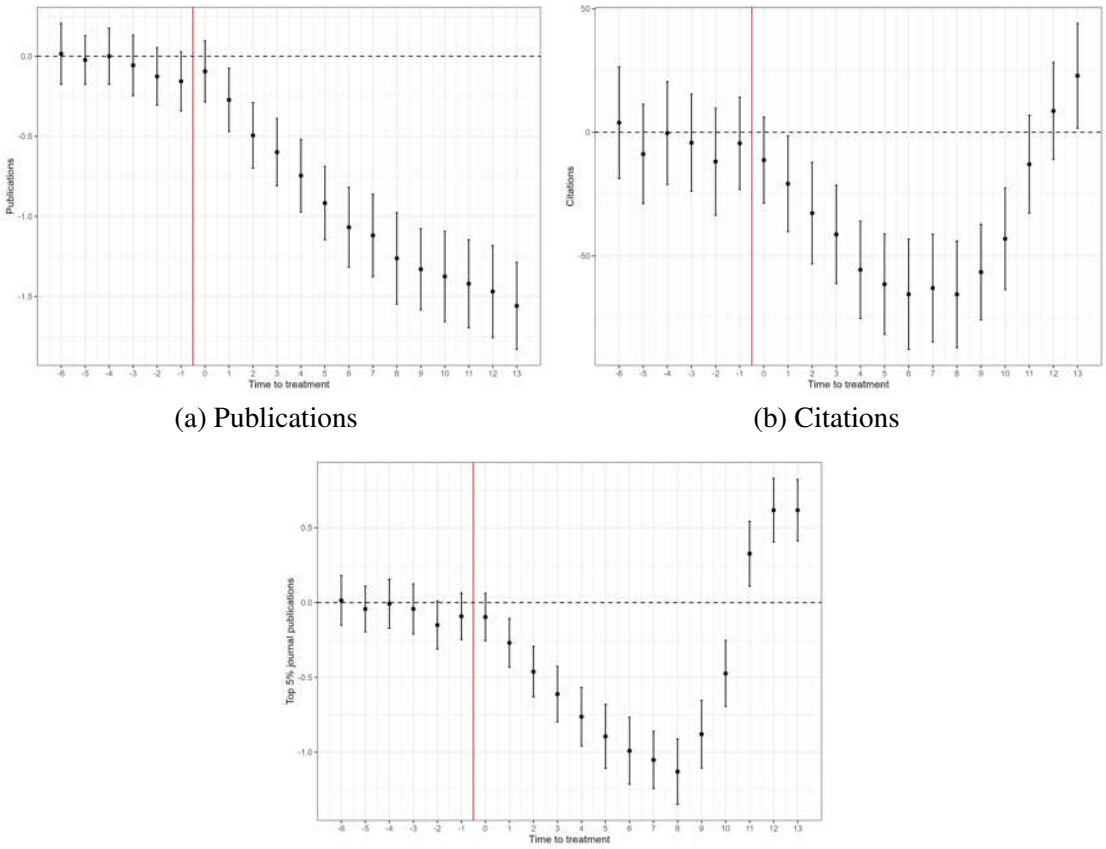
1. Did the Reform Change Researcher Productivity?

The individual-level estimation suggests that the positive total productivity effect observed at the institution-level may be driven entirely by the increase in size of institutions. Indeed, the estimated effects on publications at the individual level displayed in Figure 2.7a are very clearly increasingly negative: after 10 years, researchers in institutions treated for autonomy publish on average 1.5 fewer papers compared to the control group, around 40% of the pre-period average. Citations appear to decrease for a period and return to normal after, as shown by Figure 2.7b. For publications in the top cited journals, the reversal appears to be so strong that the effect becomes positive in the later years. Estimates are plotted in Figure 2.7c. Both of these measures are computed using citations, either at the publication or journal level: given that data collection practices in OpenAlex changed in 2020, which corresponds to 8 to 11 years after treatment depending on the cohort, some of the reversal may be a pure data artifact: estimates obtained when limiting the period of analysis to 2003-2019 are all significantly negative for all of these variables as shown in Table 2.D.1.

The same reversal is shown in Figure 2.D.1. The Arts, Melluso, and Veugelers (2025) novelty measure of new phrase combinations weighted by how often they are later reused in OpenAlex increases compared to the control group after eight periods; this translates to a small, insignificant effect on average number of such phrases in publications. Again, the reversal in trend may be partially caused by the change in data collection: the estimate for the total number is significantly negative, and not significant for the average when truncating the panel to 2019 (see Table 2.D.1). Overall, total production has decreased when only considering the number of papers, but this does not appear to hold true when considering quality, which only decreases up to around ten years before at least returning to normal, or even increasing based on post-2020 data.

Table 2.4 reports the estimates for the ATT over the entire period for these productivity outcomes. The improvement observed at the last period does not appear to outweigh the negative effects for the first part of the period. Publications drop on average by 40%, top 5% journal publications by 29%, and citations by 26%. Although the effect at the end of the period is positive, the average effect is globally negative: the cumulated effects of the reform are overall negative until 2025.

Given that the number of researchers exiting the sample increases significantly in treated institutions compared to control institutions, it could be possible that the effect on publications is caused by an increase in retirement rate. This does not appear to be the case: the effect persists



(a) Publications

(b) Citations

(c) Top 5% Journal Pub.

Figure 2.7 : Effect of Administrative Autonomy on Publications, Citations and Publications in the Top 5% Cited Journals - Individual Level

Notes: This graph reports estimates of the effect on the number of publications, citations and top 5% journal publications for researchers who were affiliated to a treated institution in 2007. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific field, five year bin of year of first publication, quartile in terms of total publications and total citations between 2000 and 2002. Variables are winsorised at the 1% level.

	Publications (1)	*Top 5% journal publications* (2)	*Citations* (3)	*Total #new phrase combinations* (4)
All	-1.565***	-0.829***	-61.167***	17.235**
Std. error	(0.118)	(0.115)	(13.462)	(8.260)
Pre-treatment mean	3.679	2.806	232.842	242.675
N (units)	66,771	66,771	66,771	66,771
Controls				
Entry year (5y. bins)	Yes	Yes	Yes	Yes
pub_n_tile	Yes	Yes	Yes	Yes
cit_n_tile	Yes	Yes	Yes	Yes

Table 2.4 : Effect of the Reform on Productivity at the Individual Level

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on total publications (1), publications in the top 5% most cited journals (2), total citations (3), total (4) and average (5) number of new phrase combinations weighted by their reuse on researchers whose affiliation in 2007 was subsequently treated. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

when restricting the sample to years in which authors have not exited academia[74]. Table 2.D.2 shows that researchers are 4.6 p.p. less likely to publish in any given year after the reform, compared to a pre-reform rate of 77%. The effect on publications is attenuated compared to the estimate above, reaching about -1.1 instead of -1.6. This however means that exit can only account for less than a third of the effect described above. This effect represents about 31% of the pre-period average of this restricted sample; given that the effect on the probability to publish reaches about 6% only of the pre-period average, the extensive margin is only responsible for about a fifth of the entire bite of the effect. In other words, while autonomy appears to push researchers towards an earlier exit from academia, the majority of the negative effect is concentrated at the intensive margin, on researchers who are publishing. Effects on citations-based quality measures remain negative, but more attenuated than effects on publications, at around -13% for publications in the top 5% cited journals and -15% for citations. Exit and a higher number of non-publishing years for affected researchers may therefore account for a share of the results, but for the most part the drop in productivity comes at the intensive margin. Estimates compared to the pre-period average fall in the same range at 29% for publications when restricting the sample to the set of researchers whose last publication in OpenAlex was recorded in 2025, which represents 48% of the sample. These results are reported in Table 2.D.3.

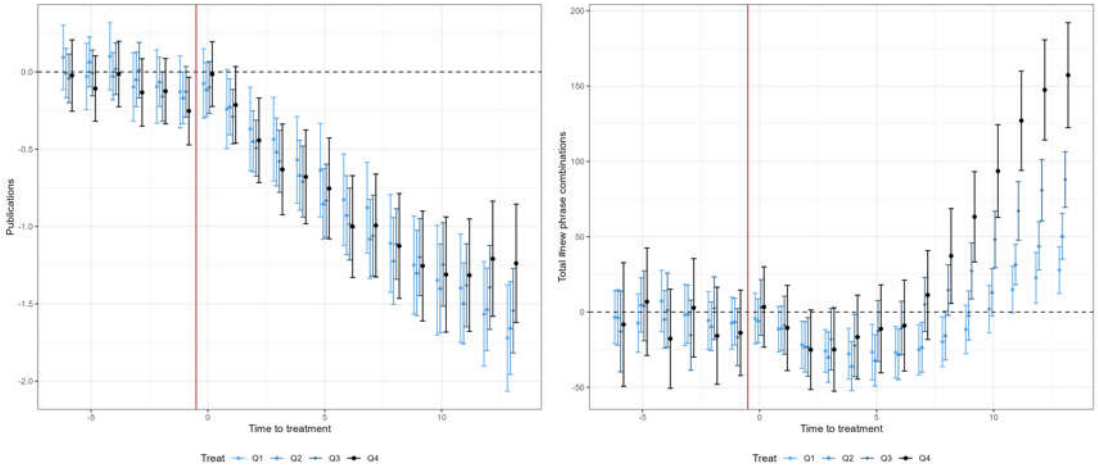
2. Who Reaps the Gains of Autonomy?

I now perform heterogeneity analysis to understand who benefits from administrative autonomy.

I first consider the same margin as above: if the university invested in IDEX-aligned governance, proxied by whether the university received an IDEX in the first call, the second call, or none, do researchers in this institution publish more on average? Figure 2.D.2 confirms the results obtained at the institution level: researchers in all three types of institutions follow the exact same trend in these separately estimated regressions, both in terms of publications and novelty. Again, this is surprising, if only because IDEX came with substantial funding opportunities, which could lead to vast differences in output if the average researcher was able to use some for her own research.

[74] Year of exit is defined as the last year in which a researcher is observed with at least one publication. This definition conditions on the outcome and may mechanically raise publication rates in the retained sample; the estimates should be read with this caveat in mind, as the ATT at each period conditional on being still active during that period.

Unlike results at the institution-level, however, pre-reform research intensity does matter at the individual level. Independent of the IDEX wave of the institution of affiliation in 2007, I estimate the same equation separately for the four quartiles of citations received for papers published during the period 2000-2002. The trends are indistinguishable regarding publications, as shown by the left panel of Figure 2.8; however, when looking at the number of new phrase combinations weighted by reuse, a pattern clearly appears. Researchers whose 2000-2002 publications attracted more citations suffer less in the first, negative part of the trend and increase their novelty significantly more than those who were cited less. Table 2.5 summarises the ATT on the same measures of productivity as above. The gradient is clear: the estimates are the same for publications, but become more positive for higher quantiles for all other outcomes, becoming clearly positive for novelty. Moreover, even though estimates are indistinguishable from each other in absolute terms, minus one publication over the period does not represent the same amount for researchers in the lowest quantile, whose pre-2007 output averages 2.3 publications, and the highest quantile, who published on average 4.6 publications per year. In other words, everyone appears affected similarly in absolute terms, but the negative effects appear to be very strong for lower quartiles, aside from novelty, while the significant negative effects for researchers with more citations to 2000-2002 papers are of a smaller magnitude. This discrepancy is consistent with reallocation of either resources or time within universities when they receive autonomy: every researcher appears to be affected, but academics who are less specialised in research are more negatively affected than others and suffer more losses over time than they reap gains.



(a) Publications (b) Novelty

Figure 2.8 : Effect of Administrative Autonomy on Publications and Novelty by Quantile of Citations to 2000-2002 Papers

Notes: This graph reports estimates of the effect on the publications and novelty proxied by total new phrase combinations weighted by their later reuse in OpenAlex for researchers who were affiliated to a treated institution in 2007. Estimates are obtained following Callaway and Sant'Anna (2021) in separate regressions based on the quantile for citations received to papers published in 2000-2002, the lighter points corresponding to the lowest quantile and the black points to the highest. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific field, five year bin of year of first publication, quartile in terms of total publications and total citations between 2000 and 2002. Variables are winsorised at the 1% level.

VI — Discussion

	Publications (1)	*Top 5% journal publications* (2)	*Citations* (3)	*Total #new phrase combinations* (4)	*Avg #new phrase combinations* (5)
Q1	-0.981***	-0.366***	-30.246***	-6.789	0.754
Std. error	(0.120)	(0.067)	(7.197)	(6.559)	(1.828)
Pre-treatment mean	2.232	0.542	40.803	46.462	14.095
N (units)	9,813	9,813	9,813	9,813	9,813
Q2	-1.044***	-0.394***	-21.865***	-0.010	1.173
Std. error	(0.102)	(0.059)	(6.441)	(6.913)	(1.420)
Pre-treatment mean	2.492	1.152	78.006	103.283	30.081
N (units)	16,459	16,459	16,459	16,459	16,459
Q3	-0.988***	-0.296***	-19.902**	23.405***	4.019***
Std. error	(0.096)	(0.071)	(7.977)	(7.731)	(1.445)
Pre-treatment mean	3.044	2.262	157.982	218.184	52.729
N (units)	16,208	16,208	16,208	16,208	16,208
Q4	-0.930***	-0.239**	-5.745	49.928***	8.138**
Std. error	(0.137)	(0.109)	(11.776)	(13.321)	(4.145)
Pre-treatment mean	4.620	3.962	312.222	382.945	73.451
N (units)	9,090	9,090	9,090	9,090	9,090
Controls					
Entry year (5y. bins)	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes
Quartile of Cit. (selection)	Yes	Yes	Yes	Yes	Yes

Table 2.5 : Effect of the Reform on Productivity at the Individual Level by Quantile of Citations to 2000-2002 Papers

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on total publications (1), total citations (2), total publications in the top 5% most cited journals (3), total (4) and average (5) number of new phrase combinations weighted by their reuse on researchers whose affiliation in 2007 was subsequently treated. Each line results from a separate estimation, using as the treated groups researchers who belong to each quantile in terms of citations to 2000-2002 papers. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

1. Robustness

I discuss in this subsection the potential problems with my identification strategy and the robustness tests I perform to ensure it holds.

Incompleteness of data: for tractability, I restrict myself to productivity and mobility within France; this prevents me from determining the direction of any brain drain, but also from taking into account potential effects of foreign affiliations on productivity of French researchers. If a researcher has an affiliation outside of France which is responsible for some of their productivity, this will not be taken into account by the statistical model. However, this increases the outcomes at the individual level; hence, the bias would be absorbed by the individual fixed-effect and not by the treatment effects. Another issue could be caused by the different levels of international competition that different academic disciplines undergo, which then also drives their recording in the data. International bibliometric databases typically under-report social sciences publications due to the fact that they are more often published in the national language. In the French context however, OpenAlex specifically retroactively includes all of the HAL repository in its sources[75], which mitigates the issue.

Control variables, selection of observations and comparison group: both sets of regressions are performed with a determined set of control variables, using never-treated and not-yet-treated units as a comparison group. I provide in Appendix 2.E.1 the unconditional pre-trends based on the same Callaway and Sant'Anna (2021) estimator without control variables and therefore without a selection equation. Universities had a choice in when they were treated, which raises an issue about whether they are comparable to each other as not-yet-treated units. In order to address this concern, I also show the time estimates for each group, as well as results for the same regressions using only never-treated units as a control group, never comparing the treated units to each other.

[75] See the 2023-12-20 release notes for the OpenAlex data dump.

Intent to Treat: in the subsection above, the analysis considers that a researcher is treated at the same time as their 2007 affiliation. The reason to choose the date is simply due to the fact that this is the year the reform passed. This is therefore the population that was targeted at the time, and looking at affiliation after the reform may be biased due to strategic moving on the part of researchers who want to avoid or join institutions that become autonomous sooner. However, results are robust to using each year from 2006 to 2009 as the year in which the researcher's university is treated, as shown in Appendix 2.E.2. On average, sample researchers in treated units stayed at a rate of 56% in their affiliation from 2007 to 2009; a further 39% have remained in an affiliation that was treated for autonomy, with over 70% of them keeping the same treatment date. 19% of control researchers had moved to treated institutions by 2009. During the last five years of the period of analysis, among researchers whose last publication was published after 2020, 93% remain in treated institutions. It thus appears that while some treated researchers have indeed left the treated institutions before or during the treatment, the vast majority of them kept being treated, with the same treatment date. Conversely, some of the controls may have been treated: this biases the estimates of this analysis towards 0 by understating the difference between control and treated units.

2. External Validity and Aggregate Implications

Unobserved variables (funding and teaching): while most of the differences entailed by the reform are related to funding, I cannot observe it until 2013. Even then, it is mostly not possible to attribute funding to a given researcher and/or institution due to the fact that affiliations tend not to be precise enough to look at the specific laboratory a researcher is working in. This is only a problem insofar as funding changes systematically in a way that is correlated with the different treatments and biases the estimates. For administrative autonomy this is not a problem: the global effect of the reform, funding changes included, is what the regression aims to measure. Funding prior to treatment is heavily restricted, so that any changes in funding after are necessarily linked to the treatment and supposed to increase productivity (i.e., private funding through university foundations). In terms of teaching, it is hard to say anything on the basis of bibliometric data. The only insights come from earlier literature. On the theoretical side, the Holmstrom and Milgrom (1991) framework would predict that the task that is less rewarded out of the two could suffer most from the change; if teaching is less evaluated and rewarded than research, it is possible that teaching suffers, even if research is less efficient due to increased effort to obtain funding to the detriment of other tasks. On the empirical side, Cantner et al. (2026) allow a more optimistic point of view: even if the grant directly targets research outcomes rather than teaching outcomes, there is no significant negative effect on the latter. Earlier literature does not allow me to make definite statements about the effect of the reform on teaching. In Section V, I find that the average production of incumbent researchers decreases more if they are in a low quartile for citations received to 2000-2002 publications. It is possible that this effect would come from specialisation. In order to become more attractive for researchers specifically, the first method would be to provide them with higher wages[76]. The second is to allow them to teach less by shifting their teaching load to researchers who publish less, which would explain the heterogeneity. However, an alternative solution would be to use autonomy to hire non-professor teachers in order to alleviate it using the aforementioned employment margin. This phenomenon is underlined by the official senatorial report tasked with assessing the effects of the reform, which shows that in 2013, 30% of all university employees were hired under a short-term, less protective contract than that of a usual civil servant (Gillot and Dupont, 2013). The fact that researchers already present in the 2000-2002 period are still affected shows that either the employment margin does not suffice, or that there are other tasks performed by those researchers.

[76] The government added to the existing bonus for "research and higher education" a new bonus of "scientific excellence" focused exclusively on research performance.

Autonomy allows universities to allocate funding more discretionarily than before; that leads them to devise means to allocate it. In particular, many of the IDEX worked as so-called "local ANR", which means that they themselves use calls for projects and contests to divide their new funding. If the 2007 reform of higher education and research in France resembles the introduction of a costly contest similar to the one developed in Gross and Bergstrom (2019), a first welfare implication of the reform is that if the contest is effective (i.e., there is some separation for higher types), regardless of what happens in the aggregate, types that are not selected for competitive funding are losing compared to the non-competitive situation. The reform should have an important distributional impact, both at an individual level and at a geographical level. Competition implies differences in outcomes; if the reform is successful, then by design good researchers should concentrate in better institutions, which means that it should increase geographical inequality in research and teaching. However, if autonomy allows productive institutions in less-favoured and less-populated regions to be more attractive, the effect on geographical inequalities could be positive. Figure 2.B.4 shows that perhaps unsurprisingly most researchers and institutions in my dataset are concentrated in Paris; however, the average number of citations is not fully correlated with this presence. There is therefore good reason based on the results to believe that the reform could have helped attract researchers outside Paris.

VII — Conclusion

In 2007, the government announced the reform of French universities as a major turning point for French academia, designed to raise productivity for the vast majority of researchers through increased autonomy in universities, allowing a competitive allocation of resources. This was put in practice from 2010 on, as a budget of 8 billion euros was allocated by a contest between clusters of universities. The evidence on the success of this strategy is mixed. This paper uses the longstanding differences in status between universities and other institutions of higher education and research to understand the effect of administrative autonomy and competitive funding.

Regarding researcher mobility, the reform appears to have met its intended goals of improving the attractiveness of universities to researchers, in particular from abroad. However, I show that for incumbents, total publications drop, and quality measures all follow a similar trend of decreasing before rising above their former level. This rebound in quality appears to be mostly driven by highly cited researchers, while less cited researchers appear to bear the brunt of the cost of reorganisation. Further research would be needed to ascertain what drives this apparent reallocation, and in particular which tasks. While teaching and school administration are likely causes, it is also possible that in this competitive environment, university academics less specialised in research also focus more on obtaining resources for more specialised researchers to use.

Chapitre 3
The Life Cycle of Ideas

Ce chapitre reprend l'article « The Life Cycle of Ideas », coécrit avec Philippe Aghion, Antonin Bergeaud, Luc Paluskiewicz et Gaétan de Rassenfosse.[77]

I — 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) and Gordon (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, Jones, 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 (i.e., Jones (1995), Segerstrom (1998), and 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, Bergeaud, et al. (forthcoming) 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, Bessen, and Wang (2025) estimate that the output elasticity (and gross marginal returns) of firm R&D in manufacturing have risen since the late 1970s, while

[77] 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.

faster obsolescence—linked to intensified technological rivalry—reduces the accumulation of effective knowledge and dampens aggregate productivity growth.[78]

We contribute to this debate by examining the scientific production of two different fields. In their structural estimation exercise based on surveys of academics, Bertolotti, Myers, and Tham (2025) show that fields strongly differ in their dependence on external funding. In their classification, economics is the field that depends the least on funding while chemistry is the one that depends on it the most. In this project, we thus study the rise and fall of topics in research in these two fields sitting at the two extremes of the distribution of funding dependency, seeking to characterize the production of ideas. We apply a machine-learning method that clusters publications based on their content, and interpret clusters as meaningful "topics" in economics and chemistry. Most of these topics behave in a similar fashion as the specific sectors in Bloom, Jones, et al. (2020) in relative terms, rising, reaching a peak in importance, then falling. The empirical analysis relies on the abstracts of over 360,000 journal articles in economics and 2.5 million journal articles in chemistry published from 1950 to 2020. Results in economics are validated using our expert knowledge, in particular to assess the meaningfulness of the topics generated by our algorithm. Specifically, we implement unsupervised, text-based cluster identification (Dieng, Ruiz, and Blei, 2020). Our method fixes the number of topics throughout the study period, though topics can be more or less active over time. Each paper relates to all topics, weighted by their relationship based on the embeddings of words present in the paper. We let the number of topics emerge endogenously based on algorithmic performance metrics, namely perplexity on held-out data. This process leads us to consider 90 topics for economics and 80 in chemistry, each corresponding to a specific object or method of study. Having allocated all the papers in our dataset to these topics, we proceed to studying the evolution of publications within each topic over the full period.

We report three main findings that are stable across these two disciplines. First, the importance of a topic within its field tends to follow a bell-shaped curve. We measure a topic's importance using the share of articles that relate to it among the total number of new articles in a given year. Topics following this typical trajectory first catch up to the yearly average production in their discipline as they gain importance, then fall below the average progression of the entire field, more rapidly for economics than for chemistry. 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 publications displays sustained growth throughout the period, without any evidence that quality declines after the peak of topic importance is reached. That is, there is no evidence that the aggregate production of economics and chemistry publications exhibits diminishing returns. Third, we find a perpetual "reset" in the production of scientific ideas, with the continual 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, and aligns with Mokyr's more optimistic view [79].

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 and Pischke (2010) use JEL codes to partition the aggregate

[78] This mechanism resonates with evidence of rising patent-output elasticities but weaker translation of patenting into firm growth (Fort et al., 2026), with findings of time-varying R&D depreciation (Li and Hall, 2020), with work documenting changes in spillovers and rivalry (Bloom, Schankerman, and Van Reenen, 2013; Lucking, Bloom, and Van Reenen, 2019), and with firm-level evidence against simple fishing out at the knowledge-production stage (Klüppel and Knott, 2023).

[79] Our topic life-cycles also connect naturally to the growing "science of science" literature documenting slowing disruptiveness and increased canonisation as fields expand (i.e., Wuchty, Jones, and Uzzi (2007), Uzzi et al. (2013), Foster, Rzhetsky, and Evans (2015), Wu, Wang, and Evans (2019), Chu and Evans (2021), and Park, Leahey, and Funk (2023)).

production of economics 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. Recently, Harris (2026) maps a subset of economic 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. Finally, Garg and Fetzer (2026) track precisely the specific evolution of causal methods in economics based on supervised classification and find similar patterns in their trajectories.

The remainder of the paper is organised as follows. Section II presents the data and outlines our method. Section III discusses the results of the clustering algorithm. Section IV presents our main findings. Section V concludes. The data acquisition and analysis pipeline is available in open access to encourage reproducibility.[80]

II — Data and Method

1. Data Acquisition

We collect data on publications and their abstracts from OpenAlex, a bibliometric database launched in 2022 that uses open-access data from Microsoft Academic Graph (MAG) (Priem, Piwowar, and Orr, 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[81]. 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, van Eck, and Waltman, 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[82]. This process does not incorporate expert knowledge and spans the entire database, resulting in articles in management and finance often being categorised as economics articles. Conversely, economics articles may be classified in other fields. For instance, mechanism design is considered a subfield within the broader field of Decision Science[83]. Because our analysis

[80] The code will be available upon publication of the article.
[81] MAG was discontinued in 2021, so that the number of available sources for OpenAlex drops after that point (Alperin et al., 2024).
[82] See the official OpenAlex topics documentation and the relevant GitHub repository.
[83] See Appendix 3.D.

critically relies on properly scoping economics and chemistry research articles and identifying topics within these fields, we will perform our own identification of economics papers and deploy a dedicated classification algorithm to identify topics.[84]

Identifying Economics and Chemistry Papers

Delineating the scope of papers that count as "economics" or "chemistry" is a surprisingly challenging task. As economists, we could consider only papers published in the Top 5 journals or the Top 20 in economics—assuming we agree on such a 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-recognised economists and chemists, such as Nobel Prize winners, have published. However, for economics, 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.

We build our list of journals from two curated lists. 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. It is only available for economics. Within that list, we keep journals that are categorised as general or specialised economics. We exclude journals categorised in "health economics" due to the large overlap with public health journals. This yields 359 journals. To that list, we add the top 300 journals in the 2025 SCImago ranking of Economics and Finance journals, keeping only economics journals and discarding finance journals.[85] This adds 54 journals not present in the CNRS list, for a total of 413 journals. Then, we restrict it 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,[86] representing 554,224 publications in OpenAlex. For chemistry, we only use the 2025 SCImago ranking. Out of their ranking of 1,035 chemistry journals, we retain the 1,009 English-language journals available in OpenAlex, which represent 6,922,831 OpenAlex publication records.

Data Preparation and Cleaning

We retain publications that have an abstract and a valid year of publication[87], leading us to consider 362,602 papers in economics and 2,496,266 papers in chemistry published between 1950 and 2020.[88]

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. (1) We remove elements from the text that are likely to bias our analysis: plain-text citations in the abstract, information about the journal of publication or JEL codes, which are sometimes appended to the abstract; (2) we perform part-of-speech tagging in order to remove stop words, keeping only adjectives, adverbs, nouns,

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

[85] The CNRS list and the SCImago list are both available online.

[86] Some journals have several identifiers in OpenAlex.

[87] This also excludes the case of abstracts that are clearly not meaningful: for instance, when the text is clearly a webscraped text indicating how to navigate the website instead of an abstract.

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

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; (3) following Fan, Doshi-Velez, and Miratrix (2017), we remove words from a hand-curated set of "corpus-specific stopwords" (such as "paper," "economics," "chemistry" ...), and country-related terms (i.e., "United States," "French," "American," etc.) to only keep meaningful keywords; (4) we stem the words in each abstract to facilitate grouping related terms. Stemming reduces a word to its root form (i.e., "development" and "developing" are both mapped to "develop").

The left panel of Figure 3.D.3 presents the numbers of publications and author profiles in our economics sample from 1950 to 2020. Both series are increasing at a high rate over that period: economics papers have grown at an annualised 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 (2009) 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 3.D.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)[89]. 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. 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 estimate an Embedded Topic Model (Dieng, Ruiz, and Blei, 2020) on a 20% training subset of the data, stratified by year, prior to applying the algorithm and split it into a training set (80% of the subset) and test set (20% of the subset) which are identical for all specifications. All specifications are trained for 100 epochs.

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 pre-determined metric. 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 3.E 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

[89] We weight each publication in this graph by the number of co-authors to account for the increasing numbers of co-authors on economics papers over time. We further discuss the gap between the sample and the rest of the authors' publication in Appendix 3.D.

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 (i.e., 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, we 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 characterised 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,

$$P(w \mid j) = \sum_{k=1}^{K} \theta_{jk} \, P(w \mid k). \tag{3.1}$$

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, Ruiz, and Blei, 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.

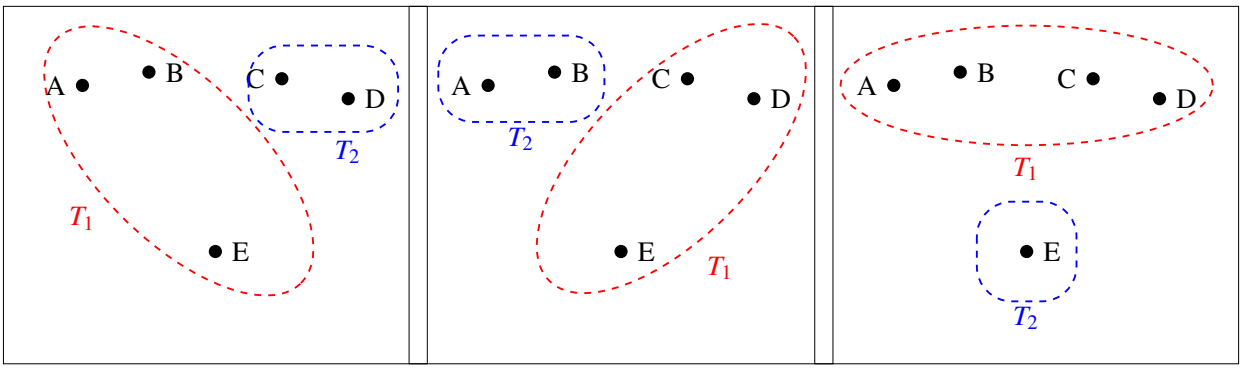
Method to Identify the Optimal Number of Topics

As an illustration, in the upper panel of Figure 3.1, 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).

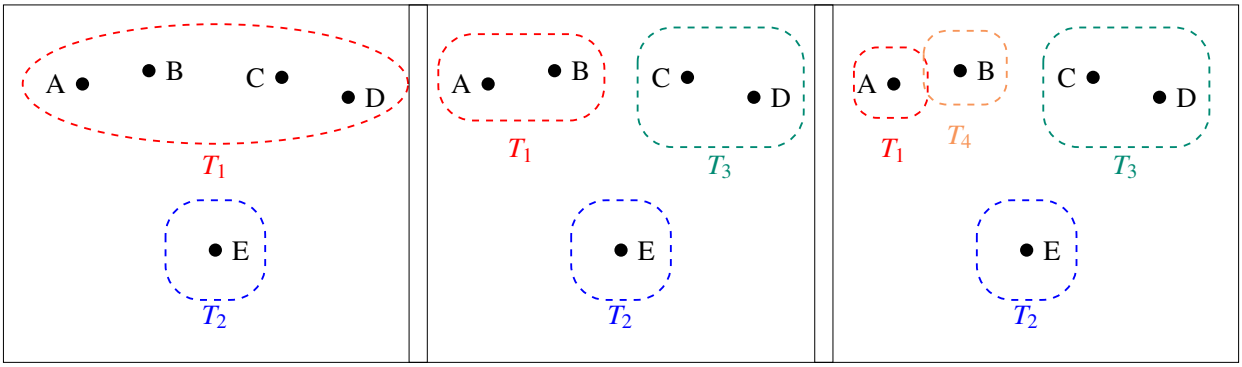
Next, we let the number of topics K vary and we assume a fixed cost per topic in order to compute the optimal number of topics K^*. We provide an illustration of this process in the lower panel of Figure 3.1. In this example, we have a trade-off between minimizing the distance between articles within 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[90].

[90] 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 3.1 : 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.

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. We measure the lack of such robustness by perplexity, which stops decreasing once 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 document-completion perplexity on the held-out test split, as implemented in the ETM package v.1.2.1: topic proportions are inferred from the first half of each test abstract, and the per-word log-likelihood of the second half is averaged across abstracts and exponentiated. 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. We provide more detailed explanations about the corresponding algorithm and validation metrics in Appendix 3.E.

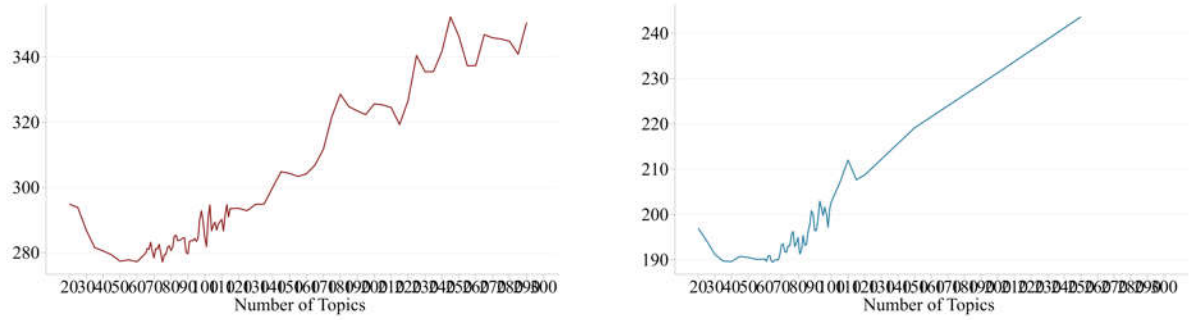
Our final algorithm uses a three-dimensional optimisation procedure that allows us to locate the kink in perplexity over a grid defined by three parameters: the minimum document frequency and the maximum document frequency of the words included in the model, and the number of topics. More specifically, we compute the values for coherence and perplexity for each of the points in the grid. We then select the model that produces a kink in the variation of the perplexity curve, determined through visual inspection. We check the robustness of

this approach by applying it to simulated data. The results of this exercise are presented in Section 3.F.2.

III — Implementation of the Algorithm

The optimisation procedure leads us to retain a minimum document frequency of 1% and a maximum document frequency of 50% (see Appendix 3.F 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 3.2 presents the evolution of perplexity in both samples. The kink in the perplexity curve is in the same region for both, although sample size for chemistry is seven times as large as economics. We set the number of topics to 90 in economics, and 80 in chemistry, but note that our analysis is robust to choosing 70 topics[91]. We then label those topics using the most associated keywords with an LLM.



(a) Economics (b) Chemistry

Figure 3.2 : Perplexity by Number of Clusters

Notes: The graph above presents, for Economics and Chemistry alike, perplexity as a function of the number of topics chosen for the algorithm on papers. 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. The model assigns a positive weight to each of the K 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 weights and then reweight the distribution based on these shares alone. 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 among their first n^{th} topics and the total number of publications in t. Again, throughout the paper, $n = 10$. We smooth this share using a weighted moving average of the current and four preceding years. Results are robust to considering the top five or top three topics. We also measure the *date of emergence* of the topic as the first year in which a topic crosses 5% of the gap between its minimum and maximum importance, before reaching its maximum.

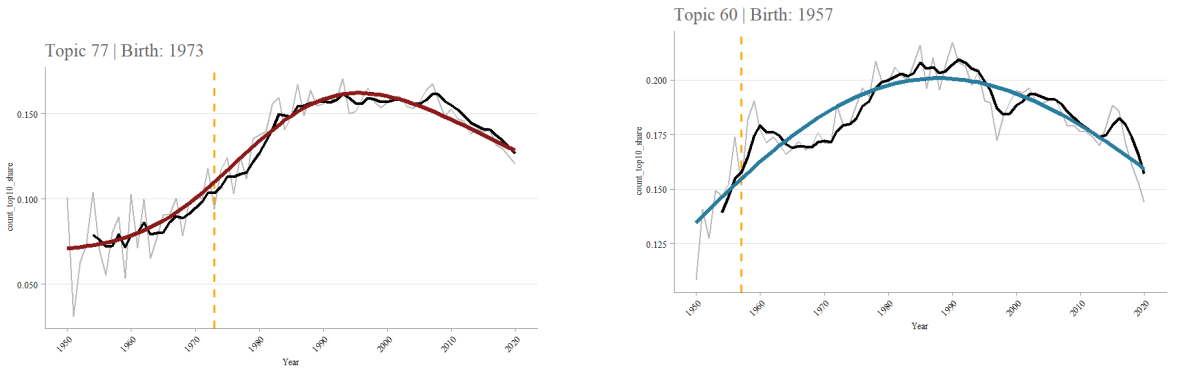
[91] Results are available upon request from the authors.

IV — Stylized Facts on the Lifecycle of Ideas in Economics and Chemistry

Having allocated papers to topics, we now seek to characterize the life cycle 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.

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

Our method does not use publication dates when identifying topics. We can therefore observe how the same topics evolve over time, using the share of publications for which topic k is among the ten highest weights θ_{jk} between 1950 and 2020. We classify trajectories by their shapes, as explained in Appendix 3.C.1: 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. About two-thirds of topics in each discipline (59 in economics and 53 in chemistry) follow an inverted-U shape after emergence: their shares rise, reach a peak and fall. This is the shape that would be expected if ideas indeed become harder to find for each new Kuhnian discovery.[92] Figure 3.3 gives an example from each discipline. Two other economics topics and one chemistry topic grow throughout the period, while five economics topics and seven chemistry topics decline throughout. Eight topics in economics and ten in chemistry follow a regular U shape. 13 topics in economics and nine in chemistry are neither unimodal nor monotonic. We are not able to assess the trajectory for three topics in economics as they concern too few papers. By our definition, 18 topics in economics and 29 in chemistry emerged before 1950. These similar patterns appear in disciplines with very different research costs.



(a) Economics: Game Theory

(b) Chemistry: Organic Synthesis – Substitution Reactions

Figure 3.3 : Topic Evolution over Time

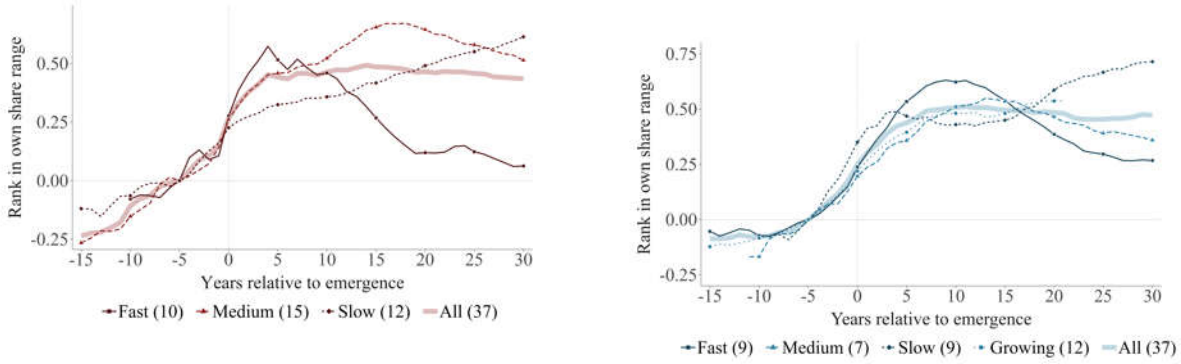
Notes: Distribution of example topics that we identify as following a bell-shaped trajectory over time. The light grey line represents the raw series, the dark grey line the series smoothed over five years, and the coloured line the fit estimated in the classification step. **a** represents the trend of a topic we label Game Theory. **b** represents the trend of a topic we label Organic Synthesis - Substitution Reactions

For the subset of topics that we identify as bell-shaped, we seek to characterize the "typical" evolution of topic importance around its date of emergence. For each topic k, let τ_k^* denote the year of emergence, and let $y_{k,t}$ be the relative rank of the value of this share compared to the minimal and maximal range, so that $y_{k,\text{peak}} = 1$ and $y_{k,\text{min}} = 0$. We define event time $t = \tau - \tau_k^*$. We then estimate the average event-time profile by running the following event-study regression:

[92] 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.

$$y_{kt} = \sum_{s=-20}^{20} \beta_s \mathbb{1}\{t = s\} + \alpha_k + \varepsilon_{kt}, \tag{3.2}$$

where y_{kt} is the (smoothed) relative rank of the share of publications that have topic k in their top 10 topics compared to the maximum at event time t.



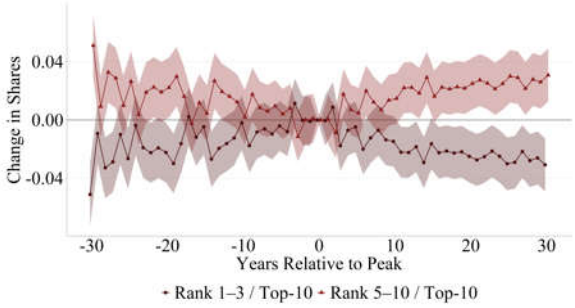
(a) Economics (b) Chemistry

Figure 3.4 : Average Topic Trends, Relative to year of emergence

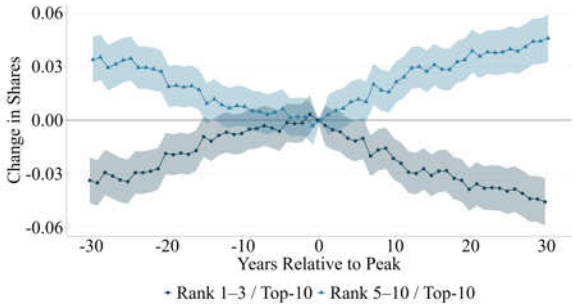
Notes: The graph above presents the average share of publications per topic relative to the year of emergence of the topic for rise-and-fall topics with a year of emergence after 1950, for economics (**a**) and chemistry (**b**), obtained by OLS, controlling for topic identifier, by duration between the date of emergence and the date of peak (Fast: lower than 15 years, Medium: higher than 15 and lower than 30, slow, higher than 30, Growing: peak in the last 5 years or indicated as growing in the classification step). The share of publications is computed as the share of publications that have this topic in the 10 highest values of their distribution per year.

Figure 3.4 plots the coefficients β_s and the 95-percent confidence intervals when estimating this equation across different groups of topics that follow a monotonic or bell-shaped trajectory, which we observe after emergence. There are 37 such topics both in chemistry and economics. Topics that we consider fast peak less than 15 years after emergence, medium topics peak between 15 and 30 years after emergence and slow topics over 30 years after emergence. We also plot the growing topics, i.e., those that are classified as such above and those that peak in the last five years of analysis. Expressing trajectories relative to the year of emergence makes them comparable to the diffusion curves documented by Griliches (1957). In both disciplines, the average line across all categories plateaus after a relatively early peak, but the split shows that this pattern is caused by composition effects, as slow and still-growing topics keep rising while faster ones are already falling. In economics, these categories are very distinct: fast topics peak about 5 years after emergence and medium topics right at the boundary of their definition at 15 years. In chemistry, with the same definition, fast and medium topics appear to follow more similar patterns: fast topics peak on average at 10 years after emergence, while medium topics peak 15 years after emergence. In both disciplines, the rise-and-fall pattern is more pronounced for papers in which a topic has one of the five highest weights θ_{jk} as seen in Figure 3.5. For papers in which it ranks between sixth and tenth, the pattern is reversed, suggesting that topics continue to appear as secondary components of publications after their relative prominence has declined.

The trajectories are consistent with the gradual exhaustion of incremental opportunities within research lines (Aghion, Bergeaud, et al., forthcoming). They could also reflect newer topics growing faster than older ones, even if the latter remain productive. Some topics rise again after an initial decline, a pattern observed for about a tenth of all topics and difficult to reconcile with a simple, irreversible process of exhaustion. More generally, the rise-and-fall pattern concerns relative publication shares. A falling share indicates growth below the disciplinary average and need not imply a fall in absolute output.



(a) Economics



(b) Chemistry

Figure 3.5 : Average Topic Trend According to Rank

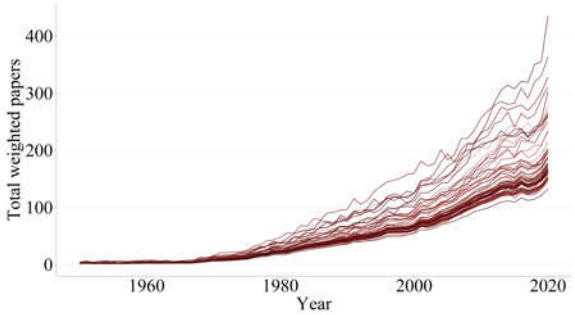
Notes: This figure depicts, for both economics (**a**) and chemistry (**b**), the average share of publications per topic relative to the year in which the maximum is obtained, normalised by the topic-specific average, for topics that belong to the 5 most associated topics to a publication (dark line), and topics with ranks 6 to 10 in a publication (clear line).

2. Global Growth in Scientific Ideas

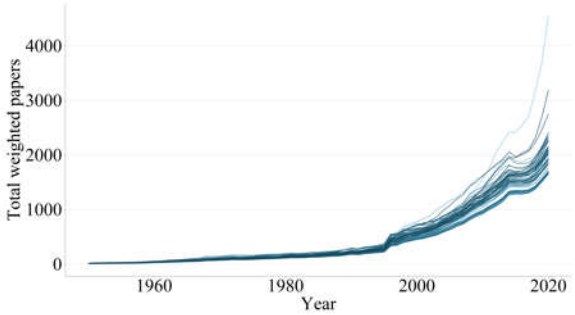
The descriptive statistics in Figure 3.D.3 depict a steady growth of publications over time. We now seek to shed light on the sources of this growth.

Topics Only Decline in Relative Importance, not in Absolute Terms

None of the topics that we identify declines in absolute publication counts. Figure 3.6 shows the number of papers within each topic, weighted by that topic's estimated share θ_{jk} in each paper. This number rises throughout the period in both disciplines. Figure 3.7 shows that average topic output, controlling for topic fixed effects, continues to increase for decades after emergence. We split topics into the same categories as those of Figure 3.4. In absolute values, the topics differ less by category in economics than in chemistry: in the former, all trajectories closely follow the average line, with fast topics growing at a smaller pace and still-growing topics at a higher pace. In chemistry, still-growing topics grow much faster, while slow-growing topics rise at a small pace compared to fast and medium topics. The increase also continues after a topic's peak in relative importance as seen in Figure 3.8. Put differently, while topics can eventually become less prominent in relative terms, the overall volume of publications per topic continues to increase.



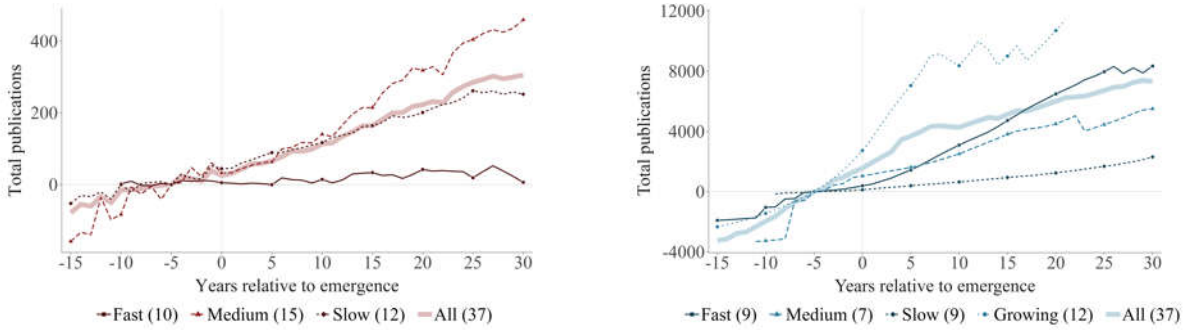
(a) Economics



(b) Chemistry

Figure 3.6 : Annual Weighted Number of Publications per Topic

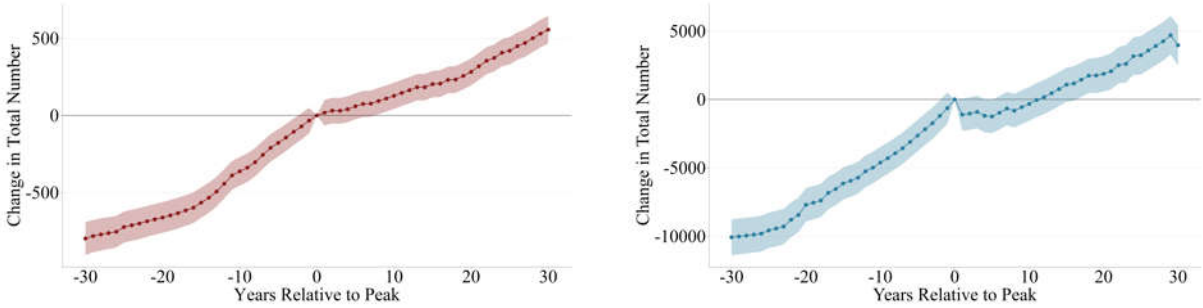
Notes: This figure depicts the evolution over time of the average number of publications per topic, restricting attention to bell-shaped classified topics, for economics (**a**) and chemistry (**b**). We count the papers for each topic, and compute the unweighted average. The weights are those estimated by the ETM model.



(a) Economics (b) Chemistry

Figure 3.7 : Average Topic Trend Relative to Emergence

Notes: The graph above presents the evolution of the number of publications per topic relative to the topic's year of emergence, for economics (**a**) and chemistry (**b**), obtained by OLS, controlling for topic identifier, by duration between the date of emergence and the date of peak (Fast: lower than 15 years, Medium: higher than 15 and lower than 30, slow, higher than 30, Growing: peak in the last 5 years or indicated as growing in the classification step).



(a) Economics (b) Chemistry

Figure 3.8 : Average Topic Trend

Notes: This figure depicts, for both economics (**a**) and chemistry (**b**), the average number of publications per topic relative to the year in which the maximum is obtained, normalised by the topic-specific average.

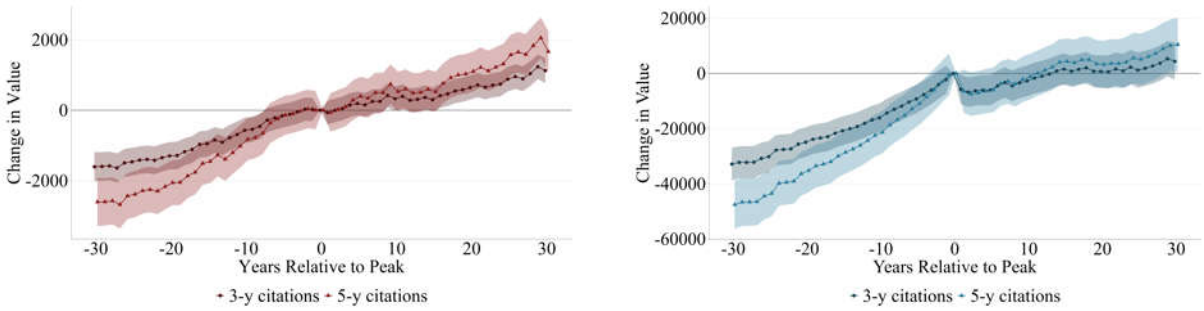
No Evidence of Quality Decrease Across a Topic's Life Cycle

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.[93] Figure 3.9 shows the average total citations received by publications in a topic in the three and five years following their publication, by year, normalised by the topic-specific average. This average number of citations received by topics increases each year. Growth 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, Jaffe, and Trajtenberg (2001), which is a measure of the dispersion of topics on which a paper builds (*i.e.*, that a paper cites), corrected for small sample bias in the Herfindahl index (Hall, 2005). We also use the generality score,

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

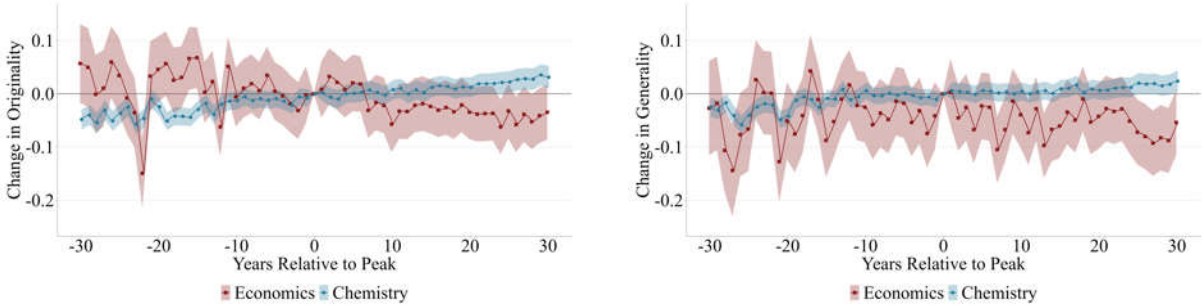


(a) Economics (b) Chemistry

Figure 3.9 : Average Topic Trend in Citations

Notes: The graph above presents the total number of citations per topic in the three (**a**) and five years (**b**) after the reference year, relative to the year in which the topic has emerged, normalised by the topic-specific average. The shaded areas represent confidence intervals for 95%.

which computes the dispersion of topics building on this paper. Figure 3.10 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.



(a) Originality (b) Generality

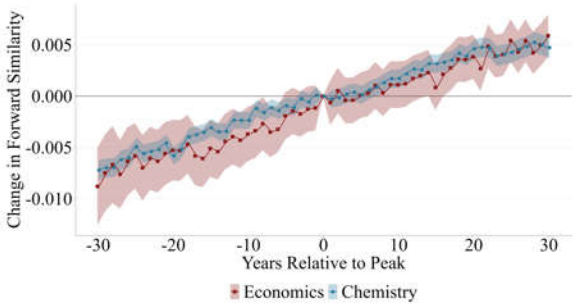
Figure 3.10 : Average Topic Trend in Originality and Generality

Notes: This figure depicts, for both economics and chemistry, the average originality (**a**) and generality (**b**) (among papers citing in the next 5 years) per topic, relative to the year in which the topic has emerged, normalised by the topic-specific average. The shaded areas represent confidence intervals for 95%.

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 that cite it. We assume that the further away the paper is located from its cited papers, the higher its novelty, in line with Arts, Melluso, and Veugelers (2025). Similarly, the closer the paper is located to the papers that cite it, the more influential it is. Figure 3.11 shows no decrease in either measure, further supporting the view that the increase in scientific output did not come at the expense of quality.[94]

[94] We further discuss measures of novelty in Appendix 3.G.

(a) Forward Similarity



(b) Backward Similarity

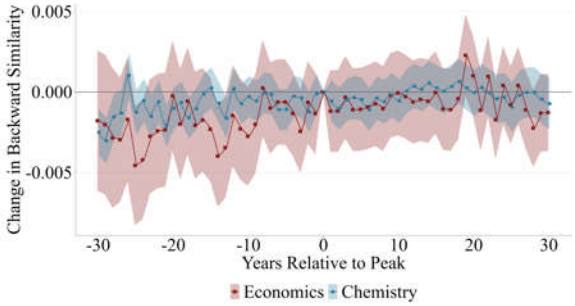
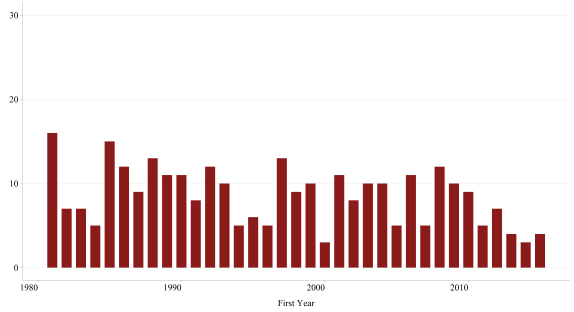


Figure 3.11 : Average Topic Trend in Similarity to Backward and Forward References

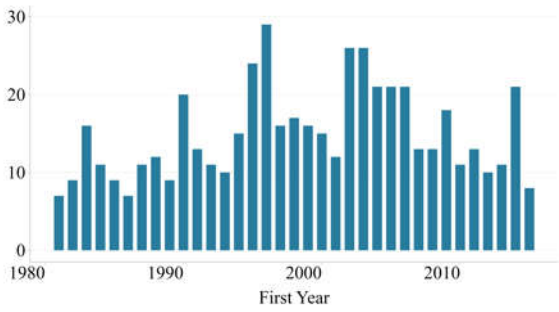
Notes: The graph above presents the average similarity to references and citing papers per topic, relative to the topic's year of emergence, obtained by OLS controlling for topic identifier. The shaded areas represent confidence intervals for 95%.

3. Reset of the Innovation Clock

Almost all topics identified by our algorithm emerge before the 1980s, possibly because the method has limited power to detect very recent topics. To assess the arrival of new research directions, we therefore examine the recombination of existing topics, a potential source of new ideas (Aghion, Bergeaud, et al., forthcoming). A "new lasting combination" occurs when two topics are the leading topics of more publications than expected under a simulated benchmark with no topical structure, for at least six consecutive years. Figure 3.12 shows that the annual number of such combinations fluctuates but remains broadly stable over 1980–2017. On average, 9 new lasting combinations appear each year in economics and 15 in chemistry. We interpret these combinations as candidates for new research directions, without assuming that each represents a radical discovery.



(a) Economics



(b) Chemistry

Figure 3.12 : Number of New Long-Lasting Combinations

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 two topics with the highest weight in each publication, independently of order. To appear in this graph, combinations of topics need to have at least six consecutive years of their share of publications exceeding a simulated threshold. This threshold is calibrated so that the annual paper count for a randomly selected pair of topics has a 5% chance of exceeding it under our noise model.

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 IV.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 3.13 shows that the topic-specific averages in our three measures of quality as computed for the aforementioned figures do not decrease on average with the year in which the topic peaks. 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 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 both disciplines, and possibly pushing up the standards for papers remaining in topics that have passed their peaks.

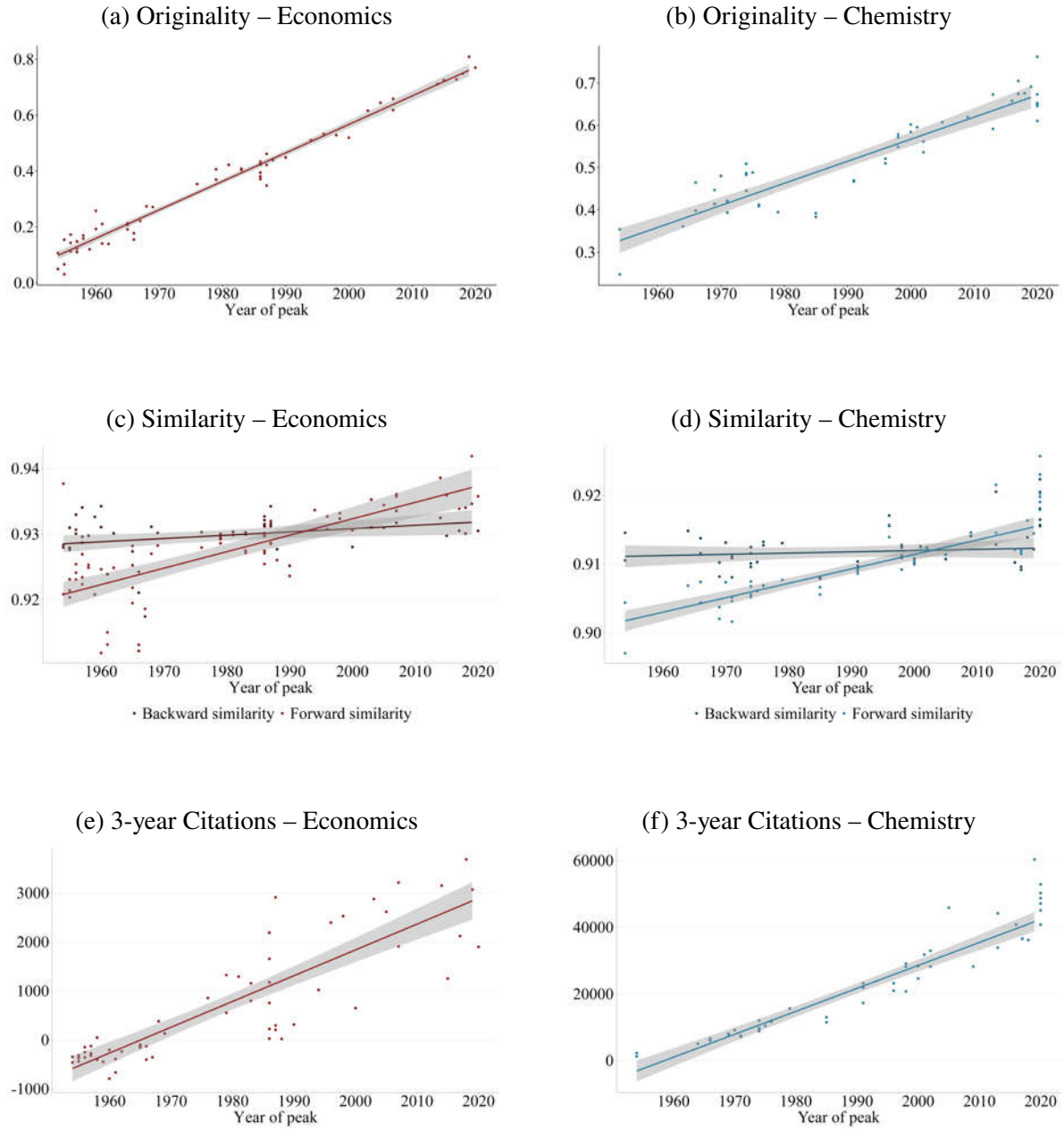


Figure 3.13 : Average Topic Trend: Originality and Similarity

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 3 years relative to the year in which the topic has emerged, normalised by the topic-specific average.

V — Conclusion

Scientific fields can renew their research agendas even as established topics lose relative importance. Our results contribute to one of the main research questions in the innovation economics literature – *are ideas harder to find?* – by showing that topic shares typically rise and fall, while publication counts within those topics continue to increase and new topic combinations keep appearing. A declining share may therefore reflect both slower growth within an established topic and the expansion of alternative research lines. This distinction is consistent with theories in which opportunities are gradually exhausted within existing lines but new lines continue to emerge (Aghion, Bergeaud, et al., forthcoming). The similar trajectories in chemistry and economics suggest that this pattern is not confined to a particular mode of research production.

Our analysis describes the evolution of scientific output but does not directly measure the productivity of research effort. More papers need not imply a proportional increase in advances, and measures based on citations and text similarity capture only some aspects of scientific quality. Similarly, new topic combinations need not correspond to radical discoveries. We therefore cannot rule out that generating new advances requires more researchers or resources over time. The continued appearance of new research directions nevertheless suggests a possible source of renewal. Whether changes in research technology, particularly computing and artificial intelligence, alter the emergence and development of these directions remains an open question.

Bibliographie

Ouvrages imprimés

Acemoglu, Daron (2009). *Introduction to Modern Economic Growth*. Princeton : Princeton University Press.

Bell, Daniel A., David Brown, Kanishka Jayasuriya et David Martin Jones (1995). *Towards Illiberal Democracy in Pacific Asia*. Springer.

Fréville, Yves (2001). *Rapport d'information sur la politique de recrutement et la gestion des universitaires et des chercheurs*. Sénat.

Gordon, Robert J. (2017). *The Rise and Fall of American Growth : The US Standard of Living since the Civil War*. Princeton University Press.

Hirschman, Albert O. (1972). *Exit, Voice, and Loyalty : Responses to Decline in Firms, Organizations, and States*. Harvard University Press.

Kuhn, Thomas S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Mokyr, Joel (2002). *The Gifts of Athena : Historical Origins of the Knowledge Economy*. Princeton University Press.

Mokyr, Joel (2017). *A Culture of Growth : The Origins of the Modern Economy*. Princeton, NJ : Princeton University Press.

North, Douglass Cecil (1990). *Institutions, Institutional Change, and Economic Performance*. The Political Economy of Institutions and Decisions. Cambridge ; New York : Cambridge University Press.

North, Douglass Cecil et Robert Paul Thomas (1973). *The Rise of the Western World : A New Economic History*. Cambridge University Press.

Popper, Karl R. (1959). *The Logic of Scientific Discovery*. English translation of *Logik der Forschung* (Vienna : Springer, 1934). London : Hutchinson.

Roberts Lyer, Kirsten, Ilyas Saliba et Janika Spannagel (2023). *University Autonomy Decline : Causes, Responses, and Implications for Academic Freedom*. Taylor & Francis.

Russell, Conrad (1993). *Academic Freedom*. London : Routledge.

Chapitre dans un ouvrage imprimé

Acemoglu, Daron, Simon Johnson et James A. Robinson (2005). « Institutions as a Fundamental Cause of Long-Run Growth ». In : *Handbook of Economic Growth*. Sous la dir. de Philippe Aghion et Steven N. Durlauf. T. 1. Elsevier. Chap. 6, p. 385-472.

Bánkuti, Miklós, Gábor Halmai et Kim Lane Scheppele (2012). « From Separation of Powers to a Government without Checks : Hungary's Old and New Constitutions ». In : *Constitution for a Disunited Nation : On Hungary's 2011 Fundamental Law*. Central European University Press, p. 237-268.

Bíró-Nagy, András (2017). « Illiberal Democracy in Hungary : The Social Background and Practical Steps of Building an Illiberal State ». In : *Illiberal Democracies in the EU : The Visegrad Group and the Risk of Disintegration*. CIDOB Editions Barcelona, p. 31-44.

Buzogány, Aron (2020). « Illiberal Democracy in Hungary : Authoritarian Diffusion or Domestic Causation ? » In : *Authoritarian Diffusion and Cooperation*. Routledge, p. 73-91.

Jenne, Erin K., András Bozóki et Péter Visnovitz (2022). « Antisemitic Tropes, Fifth-Columnism, and "Soros-Bashing" : The Curious Case of Central European University ». In : *Enemies Within : The Global Politics of Fifth Columns*. Oxford University Press, p. 45-72.

Mokyr, Joel (2014). « Secular Stagnation ? Not in Your Life ». In : *Secular Stagnation : Facts, Causes and Cures*. CEPR Press.

Pap, András L. (2017). « Constitutional Identity ? The Hungarian Model of Illiberal Democracy ». In : *A Quarter Century of Post-Communism Assessed*. Springer. Chap. 6, p. 161-186.

Rapports imprimés

Acemoglu, Daron, David Y. Yang et Jie Zhou (2021). *Power and the Direction of Research : Evidence from China's Academia*. Rapp. tech. NBER.

Adnot, Philippe (2019). *Le financement à la performance : Une condition de l'attractivité des universités françaises*. Rapport d'information 130 (2019-2020). Rapport d'information déposé le 19 novembre 2019, au nom de la commission des finances. Paris : Sénat.

Aghion, Philippe, Matthew O. Jackson, Antoine Mayerowitz et Abhijit Tagade (2023). *Innovation Networks and Business-Stealing*. Rapp. tech. 17911. CEPR.

Alabrese, Eleonora, Francesco Capozza et Prashant Garg (2024). *Politicized Scientists : Credibility Cost of Political Expression on Twitter*. Working Paper 735. CAGE.

Ando, Yoshiki, James Bessen et Xiupeng Wang (2025). *The Rising Returns to R&D : Ideas Are Not Getting Harder to Find*. Working Paper. Boston University.

Ash, Elliott, Andrea Ciccarone et Sergio Galletta (2026). *Mapping Economic Theory with AI*. Rapp. tech. 7654438. SSRN.

Azoulay, Pierre et Danielle Li (2020). *Scientific grant funding*. NBER Working Paper 26889. Cambridge, MA : National Bureau of Economic Research.

Bahar, Dany, Prithwiraj Choudhury, Sara Signorelli et James Sappenfield (2022). *Talent Flows and the Geography of Knowledge Production : Causal Evidence from Multinational Firms*. Rapp. tech. 22-047. Harvard Business School Technology & Operations Mgt. Unit.

Bai, Ying, Ruixue Jia et Jiaojiao Yang (2024). *Knowledge Suppression and Resilience under Censorship : Three-Century Book Publications in China*. Working Paper 33258. Cambridge, MA : National Bureau of Economic Research.

Becker, Sascha O., Francisco J. Pino et Jordi Vidal-Robert (2021). *Freedom of the Press ? Catholic Censorship During the Counter-Reformation*. Discussion Paper 3846201. CEPR.

Bergeaud, Antonin et Cyril Verluise (2022). *The Rise of China's Technological Power : The Perspective from Frontier Technologies*. Discussion Paper 1876. Centre for Economic Performance.

Bertolotti, Fabio, Kyle R. Myers et Wei Yang Tham (2025). *Productivity Beliefs and Efficiency in Science*. Working Paper 34000. National Bureau of Economic Research.

Bérubé, Nicolas, Gita Ghiasi, Maxime Sainte-Marie et Vincent Larivière (2020). *Wiki-Gendersort : Automatic gender detection using first names in Wikipedia*. Rapp. tech. arXiv.

Bhattacharya, Jay et Mikko Packalen (2020). *Stagnation and scientific incentives*. Rapp. tech. 26752. National Bureau of Economic Research.

Boppart, Timo et Kevin E. Staub (2016). *Online Accessibility of Scholarly Literature, and Academic Innovation*. Rapp. tech. University of Zurich.

Condorelli, Daniele (2009). *What money can't buy : allocations with priority lists, lotteries and queues*. Rapp. tech. SSRN.

Cour des comptes (2015). *L'autonomie financière des universités : une réforme à poursuivre*. Rapp. tech. Rapport public annuel 2015. Paris : Cour des comptes.

Dewitte, Edgard, Francesco Drago, Roberto Galbiati et Giulio Zanella (2022). *Science under Inquisition : The Allocation of Talent in Early Modern Europe*. Discussion Paper 17644. CEPR.

Draghi, Mario (2024). *The Future of European Competitiveness – Part A : A Competitiveness Strategy for Europe*. Rapp. tech. Brussels : European Commission.

Fan, Angela, Finale Doshi-Velez et Luke Miratrix (2017). *Prior Matters : Simple and General Methods for Evaluating and Improving Topic Quality in Topic Modeling*. Rapp. tech. 1701.03227. arXiv.

Fort, Teresa C., Nathan Goldschlag, Jack Liang, Peter K. Schott et Nikolas Zolas (2026). *Growth Is Getting Harder to Find, Not Ideas*. Working Paper 35182. National Bureau of Economic Research.

Garg, Prashant et Thiemo Fetzer (2026). *Causal Claims in Economics*. Rapp. tech. arXiv :2501.06873.

Gazzani, Andrea Giovanni, Joseba Martinez, Filippo Natoli et Paolo Surico (2026). *The public origins of American innovation*. Rapp. tech. 1521. Bank of Italy Temi di Discussione.

Gillot, D. et A. Dupont (2013). *L'autonomie des universités depuis la loi LRU : le big bang à l'heure du bilan*. Rapp. tech. Rapport du Sénat.

Hager, Sebastian, Felix Radde, Carlo Schwarz et Fabian Waldinger (2026). *Dictators, Democracies, and Discoveries : Political Institutions and the Creation of Knowledge*. Rapp. tech. 7238318. SSRN.

Hall, Bronwyn H., Adam Jaffe et Manuel Trajtenberg (2001). *The NBER Patent Citation Data File : Lessons, Insights and Methodological Tools*. Working Paper 8498. Cambridge, MA : National Bureau of Economic Research.

Harris, Tom (2026). *The Shape of Economics : Mapping the Transformation of a Discipline*. SSRN Scholarly Paper 6549201. Rochester, NY : Social Science Research Network.

Kováts, Gergely et Zoltán Rónay (2021). *Academic Freedom in Hungary*. OSUN Global Observatory Report. Central European University.

Lerner, Josh, Henry J. Manley, Carolyn Stein et Heidi L. Williams (2024). *The Wandering Scholars : Understanding the Heterogeneity of University Commercialization*. Working Paper 32069. Cambridge, MA : National Bureau of Economic Research.

Nejad, Maryam Naghsh et Andrew T. Young (2014). *Female Brain Drains and Women's Rights Gaps : A Gravity Model Analysis of Bilateral Migration Flows*. Discussion Paper 8067. IZA - Institute of Labor Economics.

Priem, Jason, Heather Piwowar et Richard Orr (2022). *OpenAlex : A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts*. Rapp. tech. 2205.01833. arXiv.

Rohner, Dominic, Oliver Vanden Eynde et Philine Widmer (2026). *Science under Threat ? A Natural Experiment in Economics*. Working Paper 2026-06. Paris School of Economics.

Wang, Liang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder et Furu Wei (2024). *Multilingual E5 Text Embeddings : A Technical Report*. Rapp. tech. 2402.05672. arXiv.

Zhao, Renyu et Yunxin Chen (2025). *Accuracy Assessment of OpenAlex and Clarivate Scholar ID with an LLM-Assisted Benchmark*. Rapp. tech. 2502.11610. arXiv.

Travaux universitaires

Corsini, Alberto (2022). « Scientific Networks and IDEX funding ». Thèse de doct. Université Côte d'Azur.

Articles de périodiques imprimés

Adda, Jérôme et Marco Ottaviani (2024). « Grantmaking, grading on a curve, and the paradox of relative evaluation in nonmarkets ». In : *The Quarterly Journal of Economics* 139.2, p. 1255-1319.

Aghion, Philippe, Céline Antonin, Luc Paluskiewicz, David Strömberg, Xueping Sun, Raphaël Wargon et Karolina Westin (2026). « Scientific Isolation ? The Consequences of Trump's China Initiative on Chinese Research ». In : *International Economic Review*.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart et Jean-Félix Brouillette (à paraître). « Resetting the Innovation Clock : Endogenous Growth through Technological Turnover ». In : *American Economic Review : Insights*.

Aghion, Philippe, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell et André Sapir (2010). « The governance and performance of universities : evidence from Europe and the US ». In : *Economic Policy* 25.61, p. 7-59.

Aghion, Philippe, Mathias Dewatripont et Jeremy C. Stein (2008). « Academic freedom, private-sector focus, and the process of innovation ». In : *The RAND Journal of Economics* 39.3, p. 617-635.

Aghion, Philippe et Peter Howitt (1992). « A Model of Growth Through Creative Destruction ». In : *Econometrica* 60.2, p. 323-351.

Aghion, Philippe et Peter Howitt (1996). « Research and Development in the Growth Process ». In : *Journal of Economic Growth* 1.1, p. 49-73.

Akcigit, Ufuk, Salomé Baslandze et Stefanie Stantcheva (2016). « Taxation and the International Mobility of Inventors ». In : *American Economic Review* 106.10, p. 2930-2981.

Alperin, Juan Pablo, Jason Portenoy, Kyle Demes et Vincent Larivière (2024). « An Analysis of the Suitability of OpenAlex for Bibliometric Analyses ». In : *CoRR*.

Angrist, Joshua D. et Jörn-Steffen Pischke (2010). « The Credibility Revolution in Empirical Economics : How Better Research Design Is Taking the Con out of Econometrics ». In : *Journal of Economic Perspectives* 24.2, p. 3-30.

Arts, Sam, Nicola Melluso et Reinhilde Veugelers (2025). « Beyond Citations : Measuring Novel Scientific Ideas and Their Impact in Publication Text ». In : *Review of Economics and Statistics*, p. 1-33.

Arts, Sam et Reinhilde Veugelers (2020). « Taste for science, academic boundary spanning, and inventive performance of scientists and engineers in industry ». In : *Industrial and Corporate Change* 29.4, p. 917-933.

Ayoubi, Charles, Michele Pezzoni et Fabiana Visentin (2021). « Does it pay to do novel science ? The selectivity patterns in science funding ». In : *Science and Public Policy* 48.5, p. 635-648.

Azoulay, Pierre, Joshua S. Graff Zivin et Gustavo Manso (2011). « Incentives and Creativity : Evidence from the Academic Life Sciences ». In : *The RAND Journal of Economics* 42.3, p. 527-554.

Azoulay, Pierre, Joshua S. Graff Zivin et Jialan Wang (2010). « Superstar Extinction ». In : *The Quarterly Journal of Economics* 125.2, p. 549-589.

Badovinac, Branka (2021). « Defining Data Quality in Bibliographic and Authority Records : A Case Study of the COBISS. SI System ». In : *Cataloging & Classification Quarterly* 59.4, p. 355-377.

Balsmeier, Benjamin et Maikel Pellens (2016). « How much does it cost to be a scientist ? » In : *The Journal of Technology Transfer* 41.3, p. 469-505.

Baron, David P. et Roger B. Myerson (1982). « Regulating a monopolist with unknown costs ». In : *Econometrica : Journal of the Econometric Society*, p. 911-930.

Batista, Catia, Daniel Han, Johannes Haushofer, Gaurav Khanna, David McKenzie, Ahmed Mushfiq Mobarak, Caroline Theoharides et Dean Yang (2025). « Brain Drain or Brain Gain ? Effects of High-Skilled International Emigration on Origin Countries ». In : *Science* 388.6749, eadr8861.

Becker, Sascha O., Volker Lindenthal, Sharun W. Mukand et Fabian Waldinger (2024). « Persecution and Escape : Professional Networks and High-Skilled Emigration from Nazi Germany ». In : *American Economic Journal : Applied Economics* 16.3, p. 1-43.

Ben-Porath, Elchanan, Eddie Dekel et Barton L. Lipman (2014). « Optimal allocation with costly verification ». In : *American Economic Review* 104.12, p. 3779-3813.

Benková, Lívia (2019). « Hungary : Orbán's Project towards 'Illiberal Democracy' ». In : *AIES Fokus* 2, p. 1-4.

Beraja, Martin, Andrew Kao, David Y. Yang et Noam Yuchtman (2023). « AI-Tocracy ». In : *The Quarterly Journal of Economics* 138.3, p. 1349-1402.

Blasutto, Fabio et David de la Croix (2023). « Catholic Censorship and the Demise of Knowledge Production in Early Modern Italy ». In : *The Economic Journal* 133.656, p. 2899-2924.

Blei, David M., Andrew Y. Ng et Michael I. Jordan (2003). « Latent Dirichlet Allocation ». In : *Journal of Machine Learning Research* 3.Jan, p. 993-1022.

Bloom, Nicholas, Charles I. Jones, John Van Reenen et Michael Webb (2020). « Are Ideas Getting Harder to Find ? » In : *American Economic Review* 110.4, p. 1104-1144.

Bloom, Nicholas, Carol Propper, Stephan Seiler et John Van Reenen (2015). « The impact of competition on management quality : evidence from public hospitals ». In : *The Review of Economic Studies* 82.2, p. 457-489.

Bloom, Nicholas, Mark Schankerman et John Van Reenen (2013). « Identifying Technology Spillovers and Product Market Rivalry ». In : *Econometrica* 81.4, p. 1347-1393.

Boda, Zsolt, Gabriella Szabó, Attila Bartha, Gergő Medve-Bálint et Zsuzsanna Vidra (2015). « Politically Driven : Mapping Political and Media Discourses of Penal Populism—The Hungarian Case ». In : *East European Politics and Societies* 29.4, p. 871-891.

Borjas, George J. et Kirk B. Doran (2012). « The Collapse of the Soviet Union and the Productivity of American Mathematicians ». In : *The Quarterly Journal of Economics* 127.3, p. 1143-1203.

Bornmann, Lutz et Rüdiger Mutz (2015). « Growth rates of modern science : A bibliometric analysis based on the number of publications and cited references ». In : *Journal of the Association for Information Science and Technology* 66.11, p. 2215-2222.

Boyack, Kevin W., Michael Patek, Lyle H. Ungar, Patrick Yoon et Richard Klavans (2014). « Classification of Individual Articles from All of Science by Research Level ». In : *Journal of Informetrics* 8.1, p. 1-12.

Buenstorf, Guido et Johannes Koenig (2020). « Interrelated funding streams in a multi-funder university system : Evidence from the German Exzellenzinitiative ». In : *Research Policy* 49.3, p. 103924.

Callaway, Brantly et Pedro H.C. Sant'Anna (2021). « Difference-in-Differences with Multiple Time Periods ». In : *Journal of Econometrics* 225.2, p. 200-230.

Camus, Jean-Yves (2022). « The Transnational Networks of the European Radical Populist Right and the Beacon of Hungarian Illiberal Democracy ». In : *Journal of Illiberalism Studies* 2.1, p. 47-54.

Cantner, Uwe, Nils Grashof, Thomas Grebel et Xijie Zhang (2026). « Rethinking excellence : the impact of the Excellence Initiative on the relative productivity of German universities ». In : *Science and Public Policy*, scag048.

Carayol, Nicolas, Emeric Henry et Marianne Lanoë (2025). « Stimulating Collaborations : Evidence from a Research Cluster Policy ». In : *The Review of Economics and Statistics* 107.4, p. 997-1009.

Card, David et Stefano DellaVigna (2013). « Nine Facts about Top Journals in Economics ». In : *Journal of Economic Literature* 51.1, p. 144-161.

Carnehl, Christoph, Marco Ottaviani et Justus Preusser (2025). « Designing scientific grants ». In : *Entrepreneurship and Innovation Policy and the Economy* 4.1, p. 139-178.

Choudhury, Prithwiraj, Ina Ganguli et Patrick Gaulé (2023). « Top Talent, Elite Colleges, and Migration : Evidence from the Indian Institutes of Technology ». In : *Journal of Development Economics* 164, p. 103120.

Chu, Johan S. G. et James A. Evans (2021). « Slowed Canonical Progress in Large Fields of Science ». In : *Proceedings of the National Academy of Sciences* 118.41, e2021636118.

Cohan, Arman, Sergey Feldman, Iz Beltagy, Doug Downey et Daniel S. Weld (2020). « SPECTER : Document-level Representation Learning Using Citation-Informed Transformers ». In : *arXiv preprint arXiv :2004.07180*.

Cologna, Viktoria, Niels G. Mede, Sebastian Berger, John Besley, Cameron Brick, Marina Joubert, Edward W. Maibach, Sabina Mihelj, Naomi Oreskes, Mike S. Schäfer et al. (2025). « Trust in Scientists and Their Role in Society across 68 Countries ». In : *Nature Human Behaviour* 9.4, p. 713-730.

Cooper, Zack, Stephen Gibbons, Simon Jones et Alistair McGuire (2011). « Does hospital competition save lives ? Evidence from the English NHS patient choice reforms ». In : *The Economic Journal* 121.554, F228-F260.

Corsini, Alberto et Michele Pezzoni (2023). « Does grant funding foster research impact ? Evidence from France ». In : *Journal of Informetrics* 17.4, p. 101448.

Cox, Gary W. et Valentin Figueroa (2025). « The Inquisition and the Decline of Science in Spain ». In : *Explorations in Economic History* 98, p. 101699.

Culbert, Jack H., Anne Hobert, Najko Jahn, Nick Haupka, Marion Schmidt, Paul Donner et Philipp Mayr (2025). « Reference Coverage Analysis of OpenAlex Compared to Web of Science and Scopus ». In : *Scientometrics* 130.4, p. 2475-2492.

Curk, Lidija, Gordana Budimir, Tomaž Seljak et Maksimiljan Gerkeš (2006). « Povezovanje Sistemov : SICRIS–COBISS. SI–Web of Science ». In : *Organizacija Znanja* 11.4, p. 230-235.

Dasgupta, Partha et Paul A. David (1994). « Toward a New Economics of Science ». In : *Research Policy* 23.5, p. 487-521.

Della Malva, Antonio, Francesco Lissoni et Patrick Llerena (2013). « Institutional change and academic patenting : French universities and the Innovation Act of 1999 ». In : *Journal of Evolutionary Economics* 23, p. 211-239.

Dieng, Adji B., Francisco J. R. Ruiz et David M. Blei (2020). « Topic Modeling in Embedding Spaces ». In : *Transactions of the Association for Computational Linguistics* 8, p. 439-453.

Dinopoulos, Elias et Peter Thompson (1998). « Schumpeterian Growth Without Scale Effects ». In : *Journal of Economic Growth* 3.4, p. 313-335.

Enyedi, Zsolt (2022). « Academic Solidarity and the Culture War in Orbán's Hungary ». In : *PS : Political Science & Politics* 55.3, p. 582-584.

Fabrizio, Kira R., Nancy L. Rose et Catherine D. Wolfram (2007). « Do markets reduce costs ? Assessing the impact of regulatory restructuring on US electric generation efficiency ». In : *American Economic Review* 97.4, p. 1250-1277.

Foster, Jacob G., Andrey Rzhetsky et James A. Evans (2015). « Tradition and Innovation in Scientists' Research Strategies ». In : *American Sociological Review* 80.5, p. 875-908.

Fukuyama, Francis (1989). « The End of History ? » In : *The National Interest* 16, p. 3-18.

Funke, Manuel, Moritz Schularick et Christoph Trebesch (2023). « Populist Leaders and the Economy ». In : *American Economic Review* 113.12, p. 3249-3288.

Gagnepain, Philippe et Marc Ivaldi (2002). « Incentive regulatory policies : the case of public transit systems in France ». In : *RAND Journal of Economics*, p. 605-629.

Gagnepain, Philippe, Marc Ivaldi et David Martimort (2013). « The cost of contract renegotiation : Evidence from the local public sector ». In : *American Economic Review* 103.6, p. 2352-2383.

Ganguli, Ina (2017). « Saving Soviet Science : The Impact of Grants When Government R&D Funding Disappears ». In : *American Economic Journal : Applied Economics* 9.2, p. 165-201.

Gaynor, Martin, Rodrigo Moreno-Serra et Carol Propper (2013). « Death by market power : reform, competition, and patient outcomes in the National Health Service ». In : *American Economic Journal : Economic Policy* 5.4, p. 134-166.

Gordon, Robert J. (2000). « Does the "New Economy" Measure up to the Great Inventions of the Past ? » In : *Journal of Economic Perspectives* 14.4, p. 49-74.

Griliches, Zvi (1957). « Hybrid Corn : An Exploration in the Economics of Technological Change ». In : *Econometrica* 25.4, p. 501-522.

Gross, Kevin et Carl T. Bergstrom (2019). « Contest models highlight inherent inefficiencies of scientific funding competitions ». In : *PLoS Biology* 17.1, e3000065.

Gush, Jason, Adam Jaffe, Victoria Larsen et Athene Laws (2018). « The effect of public funding on research output : the New Zealand Marsden Fund ». In : *New Zealand Economic Papers* 52.2, p. 227-248.

Hall, Bronwyn H. (2005). « A note on the bias in Herfindahl-type measures based on count data ». In : *Revue d'économie industrielle* 110.1, p. 149-156.

Hall, Bronwyn H. et Jacques Mairesse (2025). « Explorations of cumulative advantage using data on French physicists ». In : *Economics of Innovation and New Technology*, p. 1-20.

Halmai, Gábor (2018). « Is There Such Thing as 'Populist Constitutionalism' ? The Case of Hungary ». In : *Fudan Journal of the Humanities and Social Sciences* 11, p. 323-339.

Háló, Gergő et Márton Demeter (2023). « International vs. National Academic Bibliographies. A Comparative Analysis of Publication and Citation Patterns in Scopus, Google Scholar, and the Hungarian Scientific Bibliography ». In : *New Review of Academic Librarianship* 29.4, p. 419-438.

Hladík, Radim et Yann Renisio (2025). « Mapping Knowledge : Topic Analysis of Science Locates Researchers in Disciplinary Landscape ». In : *Poetics* 108, p. 101950.

Holl, András (2024). « MTMT - A National CRIS System in Hungary ». In : *Procedia Computer Science* 249, p. 37-39.

Holmstrom, Bengt et Paul Milgrom (1991). « Multitask principal–agent analyses : Incentive contracts, asset ownership, and job design ». In : *The Journal of Law, Economics, and Organization* 7, p. 24-52.

Hug, Sven E., Michael Ochsner et Martin P. Brändle (2017). « Citation analysis with Microsoft Academic ». In : *Scientometrics* 111.1, p. 371-378.

Iaria, Alessandro, Carlo Schwarz et Fabian Waldinger (2018). « Frontier Knowledge and Scientific Production : Evidence from the Collapse of International Science ». In : *The Quarterly Journal of Economics* 133.2, p. 927-991.

James, Chris, Lisa Colledge, Wim Meester, Norman Azoulay et Andrew Plume (2019). « CiteScore Metrics : Creating Journal Metrics from the Scopus Citation Index ». In : *Learned Publishing* 32.4, p. 367-374.

Jia, Ruixue, Margaret E. Roberts, Ye Wang et Eddie Yang (2024). « The Impact of US–China Tensions on US Science : Evidence from the NIH Investigations ». In : *Proceedings of the National Academy of Sciences* 121.19, e2301436121.

Jones, Benjamin F. (2009). « The Burden of Knowledge and the "Death of the Renaissance Man" : Is Innovation Getting Harder ? » In : *Review of Economic Studies* 76.1, p. 283-317.

Jones, Charles I. (1995). « R & D-Based Models of Economic Growth ». In : *Journal of Political Economy* 103.4, p. 759-784.

Kelemen, R. Daniel (2017). « Europe's Other Democratic Deficit : National Authoritarianism in Europe's Democratic Union ». In : *Government and Opposition* 52.2, p. 211-238.

Klüppel, Leonardo et Anne Marie Knott (2023). « Are Ideas Being Fished Out ? » In : *Research Policy* 52.2, p. 104665.

Korošec, Andrej (2022). « Prenova Aplikacije Informacijskega Sistema o Raziskovalni Dejavnosti v Sloveniji (SICRIS) ». In : *Organizacija Znanja* 27.1/2, p. 1-14.

Kováts, Gergely, András Derényi, Gabriella Keczer et Zoltán Rónay (2024). « The Role of Boards in Hungarian Public Interest Foundation Universities ». In : *Studies in Higher Education* 49.2, p. 368-381.

Krauss, Alexander (2024). « How Nobel-Prize Breakthroughs in Economics Emerge and the Field's Influential Empirical Methods ». In : *Journal of Economic Behavior & Organization* 221, p. 657-674.

Krekó, Péter et Zsolt Enyedi (2018). « Explaining Eastern Europe : Orbán's Laboratory of Illiberalism ». In : *Journal of Democracy* 29.3, p. 39-51.

Kronegger, Luka, Anuška Ferligoj et Patrick Doreian (2011). « On the Dynamics of National Scientific Systems ». In : *Quality & Quantity* 45.5, p. 989-1015.

Kwiek, Marek (2016). « The European Research Elite : A Cross-National Study of Highly Productive Academics in 11 Countries ». In : *Higher Education* 71.3, p. 379-397.

Laffont, Jean-Jacques et Jean Tirole (1986). « Using cost observation to regulate firms ». In : *Journal of Political Economy* 94.3, Part 1, p. 614-641.

Laffont, Jean-Jacques et Jean Tirole (1987). « Auctioning incentive contracts ». In : *Journal of Political Economy* 95.5, p. 921-937.

Levin, Jonathan et Steven Tadelis (2010). « Contracting for government services : Theory and evidence from US cities ». In : *The Journal of Industrial Economics* 58.3, p. 507-541.

Li, Wendy C. Y. et Bronwyn H. Hall (2020). « Depreciation of Business R&D Capital ». In : *Review of Income and Wealth* 66.1, p. 161-180.

Lucking, Brian, Nicholas Bloom et John Van Reenen (2019). « Have R&D Spillovers Declined in the 21st Century ? » In : *Fiscal Studies* 40.4, p. 561-590.

Menter, Matthias, Erik E. Lehmann et Torben Klarl (2018). « In search of excellence : A case study of the first excellence initiative of Germany ». In : *Journal of Business Economics* 88.9, p. 1105-1132.

Merton, Robert K. (1957). « Priorities in scientific discovery : A chapter in the sociology of science ». In : *American Sociological Review* 22.6, p. 635-659.

Mokyr, Joel (2018). « The Past and the Future of Innovation : Some Lessons from Economic History ». In : *Explorations in Economic History* 69, p. 13-26.

Moldovanu, Benny et Aner Sela (2001). « The optimal allocation of prizes in contests ». In : *American Economic Review* 91.3, p. 542-558.

Möller, Torger, Marion Schmidt et Stefan Hornbostel (2016). « Assessing the effects of the German Excellence Initiative with bibliometric methods ». In : *Scientometrics* 109.3, p. 2217-2239.

Moser, Petra, Alessandra Voena et Fabian Waldinger (2014). « German Jewish Émigrés and US Invention ». In : *American Economic Review* 104.10, p. 3222-3255.

Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev et Scott Stern (2016). « Of Mice and Academics : Examining the Effect of Openness on Innovation ». In : *American Economic Journal : Economic Policy* 8.1, p. 212-252.

Papaioannou, Elias et Gregorios Siourounis (2008). « Democratisation and Growth ». In : *The Economic Journal* 118.532, p. 1520-1551.

Park, Michael, Erin Leahey et Russell J. Funk (2023). « Papers and Patents Are Becoming Less Disruptive over Time ». In : *Nature* 613.7942, p. 138-144.

Pint, Ellen M. (1991). « Nationalization vs. regulation of monopolies : The effects of ownership on efficiency ». In : *Journal of Public Economics* 44.2, p. 131-164.

Prato, Marta (2025). « The Global Race for Talent : Brain Drain, Knowledge Transfer, and Growth ». In : *The Quarterly Journal of Economics* 140.1, p. 165-238.

Price, W. Nicholson (2019). « Grants ». In : *Berkeley Technology Law Journal* 34.1, p. 1-66.

Roach, Michael et Henry Sauermann (2010). « A taste for science ? PhD scientists' academic orientation and self-selection into research careers in industry ». In : *Research Policy* 39.3, p. 422-434.

Romer, Paul M. (1990). « Endogenous Technological Change ». In : *Journal of Political Economy* 98.5, Part 2, S71-S102.

Santora, Marc (2018). « George Soros-Founded University Is Forced Out of Hungary ». In : *The New York Times*.

Sasvári, Péter (2022). « Interneten Alig Láthatók Az MTA Listás Folyóiratcikkek a Bölcsész- És Társadalomtudomány Területén ». In : *Tudományos és Műszaki Tájékoztatás* 69.8.

Secen, Sefa et Ramazan Kilinç (2026). « The War on Universities : Illiberal Tactics in Turkey, Hungary, and the United States ». In : *Perspectives on Politics*, p. 1-15.

Segerstrom, Paul S. (1998). « Endogenous Growth without Scale Effects ». In : *The American Economic Review* 88.5, p. 1290-1310.

Seljak, Tomaž (1994). « COBISS—Co-Operative Online Bibliographic System and Services in Slovenia ». In : *Program* 28.3, p. 287-293.

Stephan, Paula E. (1996). « The Economics of Science ». In : *Journal of Economic Literature* 34.3, p. 1199-1235.

Stern, Scott (2004). « Do scientists pay to be scientists ? » In : *Management Science* 50.6, p. 835-853.

Szent-Iványi, Balázs et András Tétényi (2022). « Funding-Model Changes at Hungarian Universities : Greater Autonomy ? » In : *PS : Political Science & Politics* 55.3, p. 584-586.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer et Ben Jones (2013). « Atypical Combinations and Scientific Impact ». In : *Science* 342.6157, p. 468-472.

van Leeuwen, Thed N., Erik Van Wijk et Paul F. Wouters (2016). « Bibliometric Analysis of Output and Impact Based on CRIS Data : A Case Study on the Registered Output of a Dutch University ». In : *Scientometrics* 106, p. 1-16.

Veugelers, Reinhilde, Jian Wang et Paula Stephan (2025). « Do funding agencies select and enable novel research : evidence from ERC ». In : *Economics of Innovation and New Technology*, p. 1-18.

Visser, Martijn, Nees Jan van Eck et Ludo Waltman (2021). « Large-Scale Comparison of Bibliographic Data Sources : Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic ». In : *Quantitative Science Studies* 2.1, p. 20-41.

Waldinger, Fabian (2012). « Peer Effects in Science : Evidence from the Dismissal of Scientists in Nazi Germany ». In : *The Review of Economic Studies* 79.2, p. 838-861.

Waldinger, Fabian (2016). « Bombs, Brains, and Science : The Role of Human and Physical Capital for the Creation of Scientific Knowledge ». In : *The Review of Economics and Statistics* 98.5, p. 811-831.

Waltman, Ludo et Nees Jan van Eck (2012). « A New Methodology for Constructing a Publication-Level Classification System of Science ». In : *Journal of the American Society for Information Science and Technology* 63.12, p. 2378-2392.

Wolak, Frank A. (1994). « An econometric analysis of the asymmetric information, regulator-utility interaction ». In : *Annales d'Economie et de Statistique*, p. 13-69.

Wu, Lingfei, Dashun Wang et James A. Evans (2019). « Large Teams Develop and Small Teams Disrupt Science and Technology ». In : *Nature* 566.7744, p. 378-382.

Wuchty, Stefan, Benjamin F. Jones et Brian Uzzi (2007). « The Increasing Dominance of Teams in Production of Knowledge ». In : *Science* 316.5827, p. 1036-1039.

Zakaria, Fareed (1997). « The Rise of Illiberal Democracy ». In : *Foreign Affairs* 76.6, p. 22-43.

Communication dans un congrès

Beltagy, Iz, Kyle Lo et Arman Cohan (2019). « SciBERT : A Pretrained Language Model for Scientific Text ». In : *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China : Association for Computational Linguistics, p. 3613-3618.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee et Kristina Toutanova (2019). « BERT : Pre-Training of Deep Bidirectional Transformers for Language Understanding ». In : *Proceedings of the 2019 Conference of the North American Chapter of the Asso-*

ciation for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers), p. 4171-4186.

Holl, András, Gábor Makara, András Micsik et László Kovács (2014). « MTMT : The Hungarian Scientific Bibliography ». In : *Uses of Open Data within Government for Innovation and Efficiency (Share-PSI 2.0 workshop)*. Samos.

Kováts, Gergely (2015). « Recent Developments in the Autonomy and Governance of Higher Education Institutions in Hungary : The Introduction of the Chancellor System ». In : *Central European Higher Education Cooperation Conference Proceedings*. Sous la dir. de József Berács, Julia Iwinska, Gergely Kováts et Liviu Matei. Budapest, p. 26-39.

Rónay, Zoltán (2018). « Centralizations and Autonomies : The Delimitation of Education by the Hungarian Government ». In : *Bulgarian Comparative Education Society*. BCES Conference Books.

Autres sources

Hava, Nikita (2018). *Incomprehensible Numbers : We Show How Much Orbán Spent on Propaganda Last Year*. `https://hvg.hu/gazdasag/20180124_Ennyit_koltottek_el_Orbanek_tavaly_propagandara__itt_a_lista`. hvg.hu, 24 January 2018. Title translated from Hungarian. Accessed 26 April 2024.

Annexe 1
Annexes du chapitre 1
How to Silence Researchers? Evidence from Illiberal Policies in Hungary

1.A — Political Background and the State of Academia in Hungary

In this section, we provide additional resources and descriptive statistics on the state of academia in Hungary (Appendix 1.A.1) as well as on the particular case of the Central European University (Appendix 1.A.2), about which we display some descriptive statistics.

1. Descriptive Evidence of the Impact of Illiberalism on Hungarian Academia

We first present aggregate evidence and contextual information on the overall state of Hungarian research before and after 2010, marking the beginning of illiberal policy implementation.

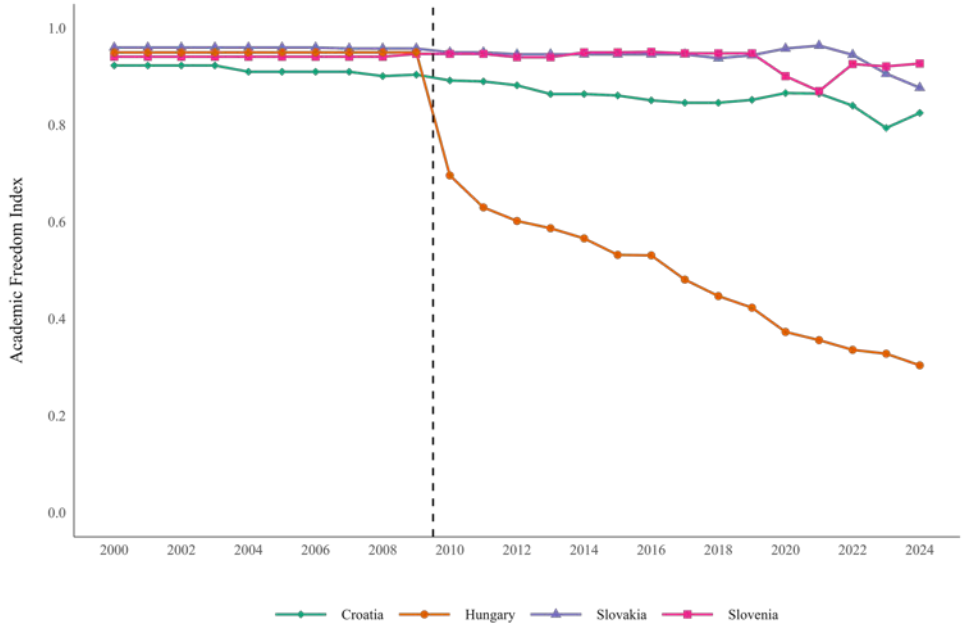
We would first like to recontextualise academic freedom in the Hungarian context by underlining the fact that the profession is used to being silent regarding political topics. The period over which academics have been involved in the public discourse without hindrance has been short, lasting only about twenty years between the end of the explicit ideological control of the communist period and the changes induced by Fidesz after 2010. During this time, academia became closely involved in politics. However, the chilling effect mentioned in this paper is a return to an earlier state from before 1989. The main difference that the literature identifies is that for those who are not self-censoring, instead of a ban on anti-government speech from academics, they are simply confined to the margins of publications by the State-aligned media, which does not even discuss their arguments. The combination of a fear for their livelihoods and the media takeover are the main reasons why academic freedom declined after 2010.

As a descriptive piece of evidence showing the rise of illiberalism in Hungary, we show that the Academic Freedom Index (AFI) has dropped since Orbán's reelection in 2010. This index is compiled by V-Dem to measure the level of academic freedom in each country on an annual basis based on country-expert surveys. In Figure 1.A.1, it is clear that Hungary diverges significantly from the trajectory of other Central and Eastern European countries[95]. The AFI for Hungary begins to decline sharply after 2010 and continues to deteriorate in the following years, while the rest only decline very slightly during Covid, which could be due to quarantine restrictions. None of the other countries experiences a drop with a magnitude similar to Hungary's. Figure 1.A.2 provides a more detailed view, displaying the individual components

[95] This is the measure that we use in order to select the countries used as a control group. The countries from which we use affiliated researchers are selected in such a way that they do not experience the same changes in academic freedom levels as Hungary.

that are aggregated to construct the overall AFI. Although all dimensions taken into account by V-Dem show a downward trend, some notable patterns emerge. In particular, the decline in "Institutional Autonomy" is especially strong in the Hungarian context, where academic freedom was initially challenged through governance reforms in 2011, 2014, and 2022. Interestingly, the indicator for "Cultural and Academic Expression" shows a sharp decline following the campaign against the Central European University and the adoption of the "Lex CEU" law. This trend may reflect the broader atmosphere of intimidation and pressure on academics between 2017 and 2019 (Kováts and Rónay, 2021; Jenne, Bozóki, and Visnovitz, 2022).

Figure 1.A.1 : Academic Freedom Index – Central-Eastern European Countries (2000-2024)



Notes: This graph shows the normalised Academic Freedom Index (AFI), computed by V-Dem, for Croatia, Hungary, Slovakia and Slovenia, between 2000 and 2024. We obtain data directly from the V-Dem website. The dashed line indicates the year Viktor Orbán took office. The value of the index goes from 0 (no Academic Freedom) to 1.

While we cannot conduct a causal analysis on the entirety of Hungarian researchers, we can still descriptively document the impact of this decrease in academic freedom.

Figure 1.A.3 highlights two notable patterns. First, it shows a slowdown in the size of PhD cohorts in public universities across all fields, with this decline becoming more pronounced after 2013. Second and interestingly, illiberal policies have not led to a reduction in the number of foreign PhD students. As the red curve in the figure indicates, the number of foreign students has continued to grow, representing an increasing share over time. In this sense, illiberal policies, in Hungary, do not appear to directly affect the arrival of foreign students; rather, their restrictive effects seem more targeted toward Hungarian researchers (see Appendix VII).

The trends observed in Figure 1.A.3 for domestic researchers are echoed in Figure 1.A.4, which displays a slowdown in the number of new researchers entering academia—especially after 2014–2015. These parallel trends also provide indirect validation of the quality of the Hungarian CRIS data used in our analysis.

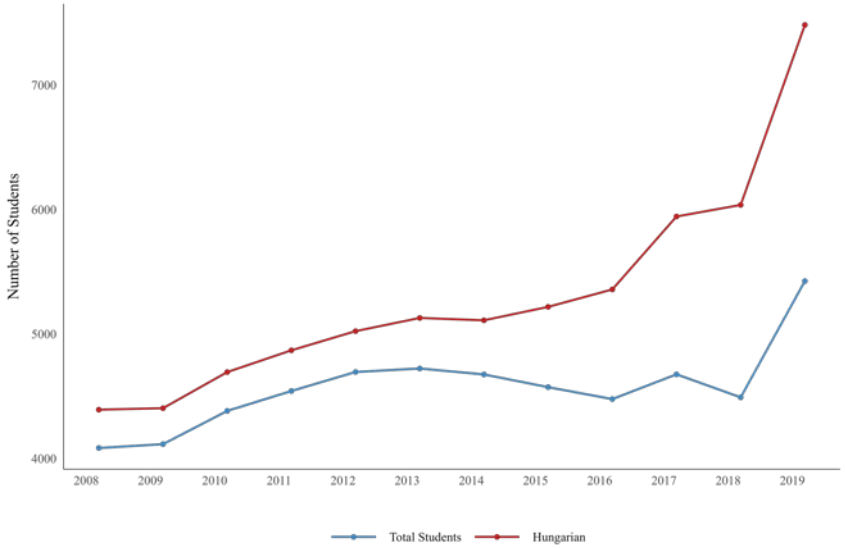
Kováts and Rónay (2021) make the case that the changes within the university and the use of media outlets to target scientists are two elements of a single strategy. This allows the government to maintain a controlled opposition within university to be able to point fingers at enemies of the regime, in a climate in which they have promoted anti-intellectualism, and to

Figure 1.A.2 : Decomposition of the Academic Freedom Index - Hungary (2000-2024)



Notes: This graph shows the decomposition of the Academic Freedom Index (AFI), computed by V-Dem, for Hungary between 2000 and 2024. We obtain data directly from the V-Dem website. The dashed lines indicate the successive 3 main organizational reforms that we describe in Appendix II as being major turning points.

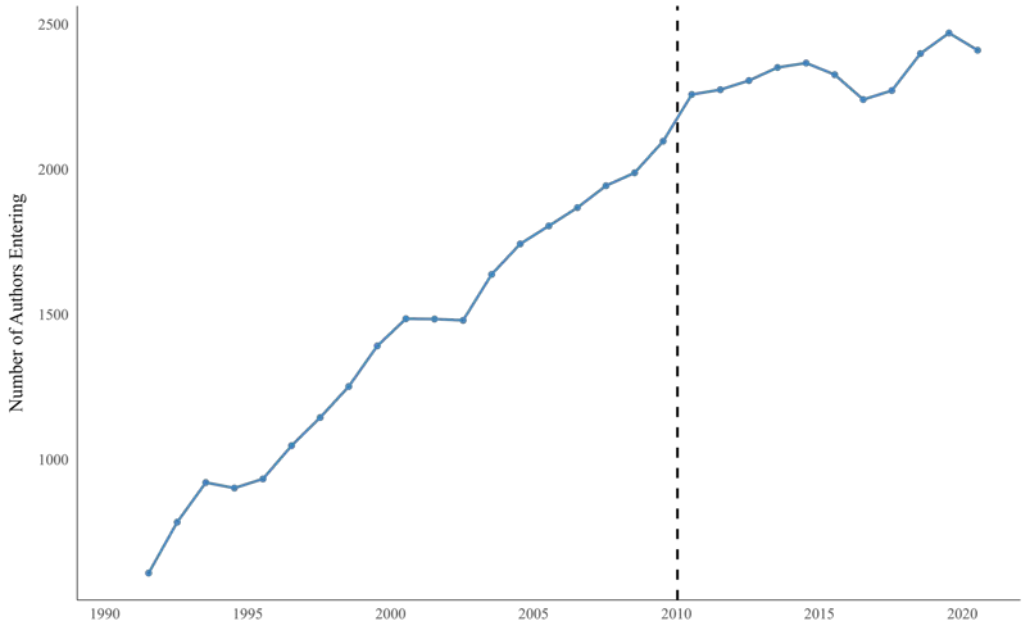
Figure 1.A.3 : Total Number of PhD Students in Hungary between 2008 and 2019 – by Nationality



Notes: The graph above displays the total number of PhD students enrolled at all Hungarian universities between 2008 and 2019. The data are sourced from the Hungarian Ministry of Higher Education. The count includes all students pursuing a doctorate, regardless of their year of study.

swiftly deal with them by controlling the internal structure of the university. While no researcher is directly told not to work on certain topics, they are all aware that they could become a target at any time. For instance, they report the following quotations from an interview with a researcher about their freedom to research and teach: "You don't need these politicians to ever have any direct say in your curriculum [...]. So they put the gun on the table and you never have to pull the trigger, and they may never pull the trigger, but that gun, you always see it. It's very powerful. In an environment that's been through that kind of socialization, it brutally restrains courage".

Figure 1.A.4 : Total Number of Entering Authors between 1990 and 2022 in Hungary



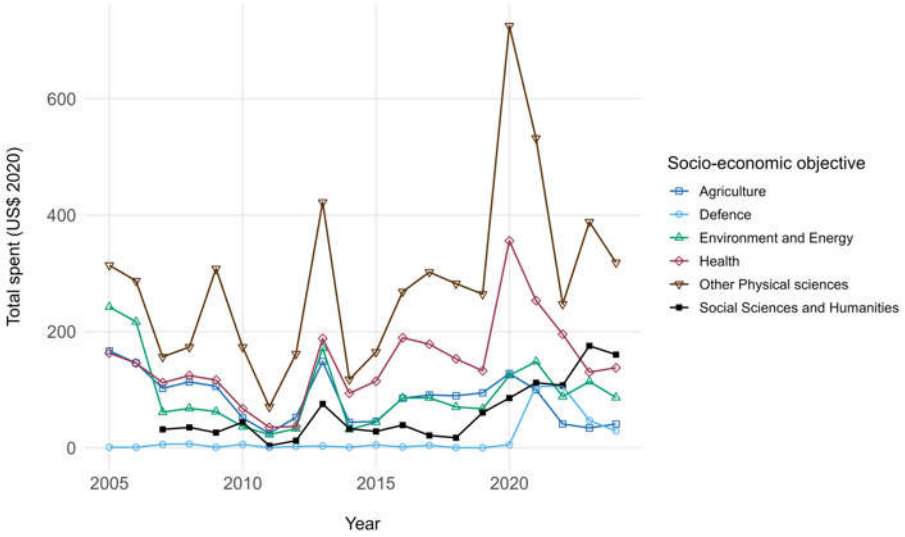
Notes: The graph above shows the total number of authors entering the Hungarian CRIS database (MTMT). An author's entry year is defined as the year of their first recorded publication. The dashed line indicates the year Viktor Orbán took office. To smooth overall trends, we compute a two-year rolling average; for example, the value shown for 2000 reflects the average number of new entries in 1999 and 2000.

Let us also remark that funding for research has not decreased overall in Hungary over the period but rather increased. Figure 1.A.5 plots the spending by large category of socio-economic objective obtained from the OECD Government Budget Allocation for Research and Development database. First, there is no great decrease in funding in any of these fields after 2010. Second, the spending in most fields appears to follow the same trends, with lower or higher spending in specific years. Third, a surprising exception in social sciences and humanities: here, the spending in Hungary clearly trends upwards. This is in our opinion a first piece of evidence that overall funding cannot drive all the negative effect of illiberalism on science.

2. The Central European University

We first recapitulate the events leading up to the relocation of the CEU to Austria. In June 2018, after the election, several "Stop Soros" legislative bills attempting to constrain the Hungarian civil society were introduced in Parliament; these laws banned people from helping undocumented immigrants and imposed a 25% tax on foreign donations to pro-immigrant NGOs (Benková, 2019). This led George Soros's Open Society foundation to move its headquarters out of the country and the EU Parliament to vote a punishment for this breach of core EU values. Benková (2019) describes the years-long campaign started in 2015 of Fidesz against George Soros, a billionaire from Hungary living in the US who is politically progressive, costing over 100 million euros. Krekó and Enyedi (2018) detail the goals as well as several stepping stones of this campaign. Finding this common enemy, the authors say, contributes to what Orbán calls a "cultural counter-revolution". The regime used its control of the media, including some private media receiving most of their funding from the state, to promote a narrative positing that progressive and pro-human rights movements in Hungary were funded by Soros, including publishing in the Figyelő a list of 200 academics and human rights activists accused of being "mercenaries of Soros". Orbán did not limit himself to such initiatives, however costly (about 250 million in US dollars on anti-Soros and anti-Brussels campaigns in 2017 alone according

Figure 1.A.5 : Total spent by the government on research by type of socio-economic objective

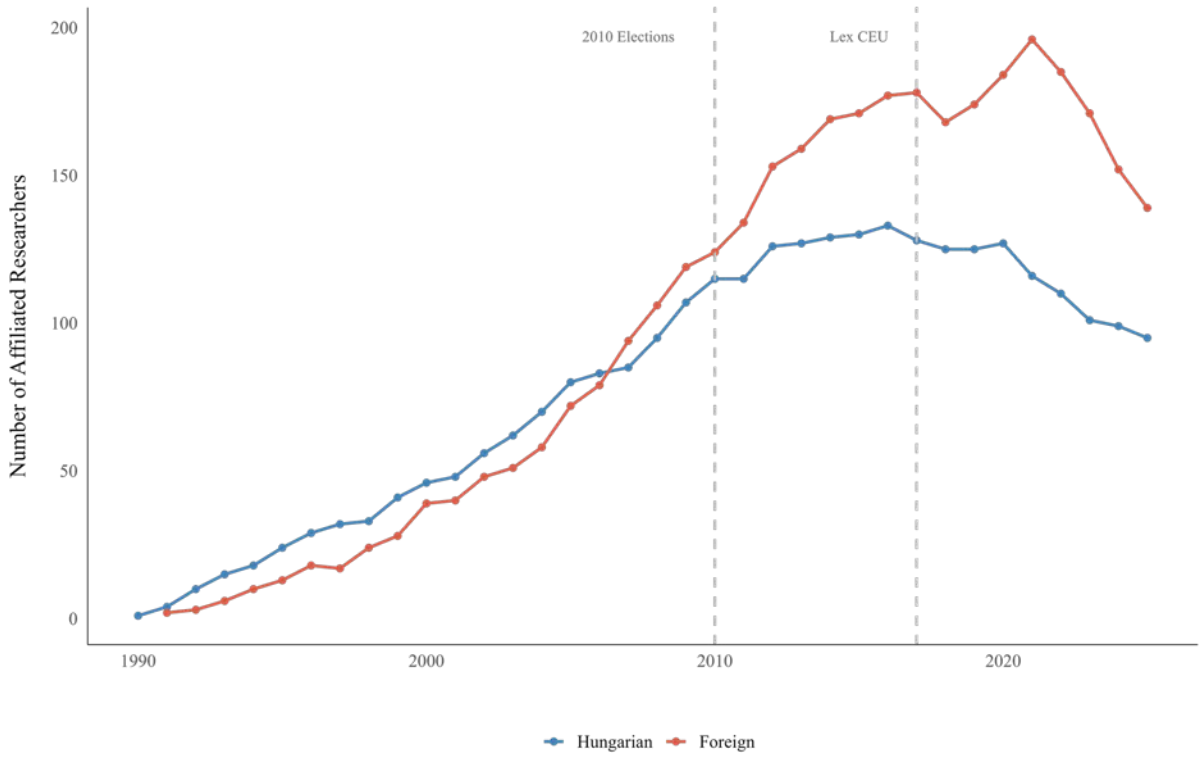


Notes: This graph shows the total amount spent in 2020 PPP US$ by the government by type of socio-economic objective between 2005 and 2024, retrieved from OECD GBARD data on 12.10.2026 on the OECD Data Explorer. The categories, based on socio-economic objectives in the database, are: agriculture, defence, environment and energy (sum of the two), health, other physical sciences (sum of the following initial categories: exploration and exploitation of the Earth, industrial production and technology, exploration and exploitation of space, transport, telecommunication and other infrastructures).

to HVG (Hava, 2018), in publications, posters...).

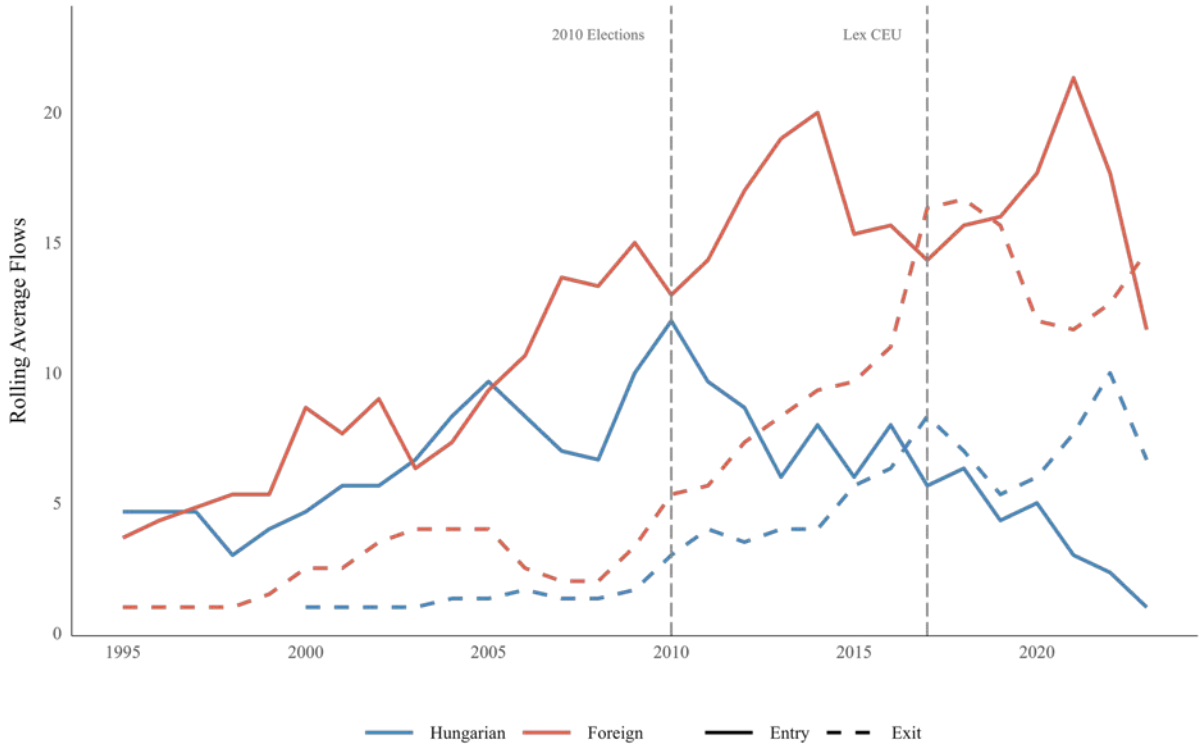
We also present in this section some general descriptive evidence on the situation of the Central European University. Figure 1.A.6 displays the number of researchers affiliated with Central European University each year between 1990 and 2025. The figure shows a steady increase in the total number of researchers—both Hungarian and international—until a shift occurs after 2010, particularly in the total number of Hungarian researchers affiliated to the university. Figure 1.A.7 provides further insight into these trends. It shows that this change is driven by both an increase in the number of researchers leaving the Central European University and a sharp decline in the number of Hungarian researchers entering. Similar trends are observed for international researchers, although the decline in entries becomes most pronounced after 2014. Trends shift again after the "Lex CEU" and the main campus of the university leaving Budapest for Vienna in 2019.

Figure 1.A.6 : Number of Professors at the Central European University (1990-2025)



Notes: The graph above displays the total number of publishing researchers that we identified as being at the CEU. This statistics only includes researchers that published at least once.

Figure 1.A.7 : Exit flows of professors from the Central European University (2000-2025)



Notes: The graph above displays the flows of entering and existing researchers at the Central European University. Data comes from our data set matched by hand. To smooth overall trends, we compute a two-year rolling average; for example, the value shown for 2000 entries is the average number of new entries in 1999 and 2000.

1.B — Data Collection and Harmonization Process

This appendix describes the construction of our dataset, which combines the OpenAlex bibliographic database with the national Current Research Information Systems (CRIS) in different Central European countries. We collected the national CRIS data for five countries: Hungary, Slovenia, Croatia, Slovakia and Czechia. However, the Slovak (SK CRIS) and Czech (IsVaVai) databases are built at the project level and suffer from inconsistencies in the registration of authors and publication dates. They are thus not yet usable in their current state, and we leave them for future work. Our analysis therefore relies on the data for the three remaining countries: Hungary, Slovenia and Croatia. We first describe how each data source was collected, then how we harmonise fields of research across sources, and finally how we merge the national and international data at the publication and author levels.

1. Collection of the Data Sets

OpenAlex is a fully open-access bibliographic database. We use the full snapshot released on 29 August 2024. This release follows a major revision of author disambiguation in OpenAlex, which led to the reassignment of all author identifiers. We still find several instances of duplicated or mixed-up author profiles when comparing them with national databases, an issue that is unavoidable given the size of the dataset.

The national repositories (Current Research Information Systems, CRIS) combine different types of sources: bibliometric data and doctoral-school records (MTMT and ODT for Hungary), project-based open data on research output (SK CRIS for Slovakia and IsVaVai for Czechia), national library records linked to a project-based researcher registry (SICRIS and COBISS for Slovenia), and a national CRIS built on existing institutional databases (CRORIS for Croatia). All are public databases, but apart from IsVaVai, none of them provides a bulk download. We therefore collect the data automatically through their API or website (MTMT, ODT, SK CRIS and CRORIS) and through the official register (SICRIS and COBISS). Although we also collected IsVaVai and SK CRIS, we cannot use them because of inconsistencies in the registration of authors and dates. Our data collection process is similar for all databases. We first collect the identifiers of all authors in the database. We then use these identifiers to obtain information at the author level, and then information about the publications of each author. This allows us to build a publication-author level dataset, which we later aggregate at the author level to construct the sample. Since all databases share the CRIS design, their variables are similar: we harmonise them and append the national datasets to build the full CRIS dataset. The harmonisation of fields of research is described in Appendix 1.B.2.

Hungary: We use two databases to characterize research activity and researchers in Hungary. The first is the list of doctoral programme members and students publicly available on the official website of the National Doctoral Council (*Országos Doktori Tanács*, hereafter ODT). Under Act CXXXIX of 2005 on higher education, all doctoral schools have been under the supervision of the ODT (an independent body) since 2007. Since November 2007, the ODT has required every doctoral school to publish a public profile for each of its researchers. These profiles include all academic titles of the researcher, the PhD students she supervises, her research areas, her employment history and a list of her recent publications. All the information is entered by researchers themselves and validated by the administrators of the doctoral school. Overall, the website contains about 52,000 profiles of researchers and PhD students. Any researcher who was active in a doctoral school after 2007 should have a profile, including now retired or deceased researchers. This information was obtained from the website doktori.hu.

The other source is the Library of Hungarian Scientific Works (*Magyar Tudományos Művek Tára*, hereafter MTMT), a publication-level bibliometric database launched in 2010. Five institutions contributed to its creation: the ODT, the Hungarian Accreditation Committee,

the Hungarian Rectors' Conference, the Hungarian Scientific Research Fund (OTKA) and the Hungarian Academy of Sciences, which also manages it (Holl et al., 2014). Records are entered and validated at the institutional level. Hungarian publications in social sciences are likely to be poorly indexed in non-specialised databases (few have a DOI, and few are indexed in Scopus (Sasvári, 2022)).[96] We retrieve the database through calls to its API using researcher profiles. We harvest 2,491,849 scientific publications out of about 5 million records[97] and 70,401 author profiles, which is consistent with the figures recently reported by the curators of the data (Holl, 2024).[98] Among those, 15,643 are directly linked to a researcher profile in ODT.

Slovenia: SICRIS is developed and maintained by IZUM, the Institute of Information Science, and by the Slovenian Research and Innovation Agency (ARIS), instituted by the Research and Development Act 2002/26 (Curk et al., 2006). The ARIS evaluates research projects and publishes the resulting data not only at the project level but also at the level of researchers, research groups and research institutes. We use SICRIS to retrieve information about active researchers in Slovenia, which represents 48,022 records overall (Korošec, 2022).

We then use the link provided by SICRIS to retrieve publication information for all these researchers in COBISS.net, the shared library information system of Slovenia, North Macedonia, Serbia, Bulgaria, Albania, Bosnia and Herzegovina, Montenegro and Kosovo. Originating in a shared cataloguing system of former Yugoslavia developed at the initiative of IZUM (Seljak, 1994), the database contains records on all publications by authors from the country, with relatively high quality (Badovinac, 2021). The COBISS.SI database contains over 7.3 million records. We only collect the subset obtained through reports generated for authors in the SICRIS database, which represents 1,515,089 records. Some COBISS entries include citations obtained from other sources, such as Web of Science, but not systematically.

Croatia: CRORIS is maintained by the Ministry of Science and Education of Croatia and the University Computing Centre. It was developed under project KK.01.1.1.03.0001 of the European Regional Development Fund (2017). The project builds upon existing databases to create and maintain a Croatian CRIS that contains 920,921 publications and 52,169 researchers. We collect the full set of CRORIS publications through its API. CRORIS does not include any citation data.

2. Harmonization of Fields

We harmonise domains and fields using the OpenAlex classification, which is itself adapted from the Scopus subject areas. This classification has 4 main domains (Life Sciences, Social Sciences, Physical Sciences and Health Sciences) and 26 fields in total, and 252 subfields.

To derive the field of an author from the national databases, we map each field of science named in the different national research repositories to the OpenAlex classification, first by hand for fields that have an identifier, then by selecting the closest OpenAlex field in terms of the cosine distance between SciBERT embeddings (Beltagy, Lo, and Cohan, 2019). We proceed in the same way to identify subfields in MTMT and SICRIS when fields are detailed enough. For instance, "linguistics" is mapped to OpenAlex subfield #1203, "Language and linguistics". We then use this mapping to retrieve the domain of publication of the author in the OpenAlex classification. Because this information is provided directly by universities or researchers themselves, we consider it more trustworthy than the automatically computed

[96] Publications in other languages are frequent in the Hungarian CRIS data, although this depends on the discipline. For instance, theology and linguistics are often published in German, less so in the Slovenian COBISS.

[97] MTMT also records non-academic publications and other types of works, for instance musical performances, which are not relevant to our analysis.

[98] A few researcher profiles are not visible to all users, but this concerns about a hundred profiles in the entire database.

OpenAlex classification. Therefore, for authors present in both sources, we use the field declared in the national database instead of the OpenAlex variable "topic share". This reduces measurement error in authors' fields.

For authors without a field in a national database, we rely on the OpenAlex variable "topic share", which gives the share of an author's publications associated with each research topic. We assign to each author the topic with the highest share, as well as any other topic whose share is at least 75% of this maximum. We then keep the subfields and fields to which these topics belong in the OpenAlex classification. We allow an author to be in several fields or domains at the same time; in that case, the author is assigned to a unique combined field corresponding to the intersection of the fields in which she publishes.

3. Merging National and International Data

In order to use the different data sources jointly and reduce measurement error in our analysis, we reconcile the information available in these databases by merging them on common identifiers. For instance, the professor of literature corresponding to ODT profile #14 has not one but five different OpenAlex profiles. One of them contains most of his publications (55), while 8 others are scattered across the remaining four. In order to compute our measures of productivity, activity and location more precisely, we use the information available in the national CRIS to identify these profiles as belonging to the same person. We detail below how we deal with the different issues arising during this process.

a) Matching Publications

In theory, merging publications across databases should be straightforward: most publications available online have a Digital Object Identifier (DOI), which should be the same across all databases. However, for most publications in our databases, the DOI is not available (in MTMT, only 18% of publications have one). Other identifiers are available, such as Web of Science or Scopus identifiers, but merging only on these external identifiers misses most of the common information across databases. Therefore, we also merge on publication title. To capture matches in which Unicode encoding, capitalisation or stopwords differ, we normalise the titles by removing stopwords when the language of the publication is known, removing non-alphanumeric characters, and converting the strings to the same encoding.

When we use matched authors to match publications, we further allow for some variation in publication titles. Stopwords may have been removed in one record but not in the other when the language is missing, and titles may contain typing errors (for instance, the title of publication #12044 in MTMT includes "labortory" instead of "laboratory", preventing a direct match with OpenAlex publication #1983568155). We therefore keep matches in which the Levenshtein distance between normalised titles is below 10% of the length of the shorter of the two strings.

b) Matching Authors

We first match authors through their publications. We keep matches between authors listed by both databases as authors of a matched publication when their first and last names are concordant, or when the OpenAlex profile only has an initial that is consistent with the first name. This allows us to match researchers even if they share the same name and field, provided that their publications are not mixed together in OpenAlex.

We then match researcher profiles on external identifiers, names and fields. We normalize the names of researcher profiles in each database and match authors on their initials and last name. We keep a match when the researcher is the only one in the national database to have this full name, or initial and last name, in a given OpenAlex field.

A researcher from a national database may thus be linked to several OpenAlex profiles. Since OpenAlex tends to create several profiles for the same author, and since we consider the national author profiles to be more accurate, we assume that when an author has no homonym in the same field in the national database, homonymous OpenAlex profiles in that field are additional profiles of the same person. This approach complements the publication-based matching, as it retrieves researcher profiles whose publications have been mixed with those of homonyms from another field. Overall, we link 16,401 Slovenian profiles and 59,629 Hungarian profiles to OpenAlex.

c) **Eliminating Incorrect Matches**

Because of homonyms in author names and publication titles, we need to eliminate incorrect matches. In principle, our procedure distinguishes authors who have the same name using their publications, but this does not work for all of them. Given the size of the dataset, the author attribution algorithm of OpenAlex generates false links between authors and publications on the basis of homonymy. For instance, some of the publications of the linguist and sociologist A5080416716 are attributed to her homonym A5004379723, who works on pulmonary diseases and endocrinology. Therefore, our process necessarily generates some false matches between author profiles.

Conversely, as explained above, we want to keep some profiles or publications that are not matched one-to-one across databases. Given the duplicated publications and author profiles, it is efficient to keep at least some of them. In this step, we consider the national databases as the source of truth, since they are at least partly managed by universities and researchers, who have an incentive to check and correct their profiles. By contrast, given the sheer size of OpenAlex, author profiles and duplicate publications cannot be checked for the entire world. Therefore, we keep all OpenAlex profiles that are matched to a single identifier in the national database, and discard those matched to several national identifiers.

Table 1.B.1 provides a summary of the results of the matching process for Hungary (MTMT) and one of the countries we use as control in the cross-country analysis, Slovenia (COBISS). Overall, Panel A shows that matched publications represent only 17.2% of total publications in MTMT and 6.2% in COBISS. However, this subset allows us to match 25.9% of researchers in MTMT and 15.3% in COBISS, as displayed in Panel B. Our next matching step further allows us to match as many as 84.7% of researcher profiles in MTMT to at least one OpenAlex profile. For COBISS, this share is 42.4%. This gap between the two data sources likely reflects the lack of external identifiers in COBISS compared to MTMT, as COBISS does not provide the DOI for each publication and only provides the Web of Science and Scopus identifiers (and only for about 10% of all publications). Conversely, in MTMT, 21% of publications have at least a DOI, Scopus, Web of Science, or PubMed identifier. Likewise, authors in COBISS do not have any external identifiers. This makes the merging process less likely to result in one match. By contrast, 26% of researchers in MTMT are also recorded with an ORCID or Scopus profile. Panel C presents the results of the matching algorithm by main domain of research of MTMT and COBISS authors.[99] Unsurprisingly, the match rate is lower for researchers in social sciences in MTMT. This is because their publications or field are less likely to be recorded accurately in OpenAlex. In COBISS, health scientists appear to be particularly well recorded in OpenAlex compared to scientists in other domains. For these other domains, match rates are similar to one another.

[99] In this computation, we count researchers once per domain of activity, i.e., a researcher active in Life Sciences and Health Sciences counts once in each of the corresponding lines.

Table 1.B.1 : Statistics on the Merging Procedure between OpenAlex and National Repositories

	Hungary (MTMT)		Slovenia (COBISS)	
Panel A: Publications				
	N	%	N	%
Total	2,491,849		1,638,784	
Step 1	316,089	12.7	52,386	3.2
Step 2	428,162	17.2	101,361	6.2
Step 3	428,462	17.2	101,373	6.2
Step 4	428,472	17.2	101,376	6.2
Panel B: Authors				
	N	%	N	%
Total	70,401		38,673	
Step 1	18,005	25.6	5,924	15.3
Step 2	18,251	25.9	5,924	15.3
Step 3	18,253	25.9	5,924	15.3
Step 4	18,254	25.9	5,924	15.3
Matched	59,629	84.7	16,401	42.4

Panel C: Authors by OpenAlex domain

	Total	Matched	%	Total	Matched	%
Life Sciences	12,089	6,556	54.2	2,365	1,369	57.9
Social Sciences	34,991	11,787	33.7	7,676	3,863	50.3
Physical Sciences	23,569	12,328	52.3	16,841	7,990	47.4
Health Sciences	12,634	7,543	59.7	3,681	2,757	74.9

Note: This table represents the number of identifiers which have at least one credible match in OpenAlex after each step of the iterative merging process, for MTMT and COBISS, for publications (Panel A), researcher profiles (Panel B) and researcher profile by main domain of study (Panel C). Next to each number, we print the percentage it represents compared to the total number of records in the CRIS database. The "Matched" step is the number obtained after iterative matching and subsequent matching on fields and names. In Panel C, Total is the number of researcher profiles in the domain, Matched is the number with at least one credible match in OpenAlex, and the percentage is Matched divided by Total.

4. Comparing Academic Trajectories across Central European Countries

This section compares publication patterns across the three national databases used in the cross-country author-level analysis, namely MTMT for Hungary, SICRIS/COBISS for Slovenia and CROSBI/CRORIS for Croatia, and checks whether differences in data collection could bias our results.

Because the databases we use were built separately, they likely started collecting publications systematically at different times. This could bias our estimates of the effect of the regime change in Hungary. In particular, funding cuts during the European sovereign debt crisis could have hindered data collection in one of the databases. To make the three databases comparable, we restrict publications to journal articles, books, book chapters and conference papers and harmonise document types across databases. In all different data bases, once a researcher's account is created, publications are updated automatically, so recorded output does not depend on individual reporting effort.

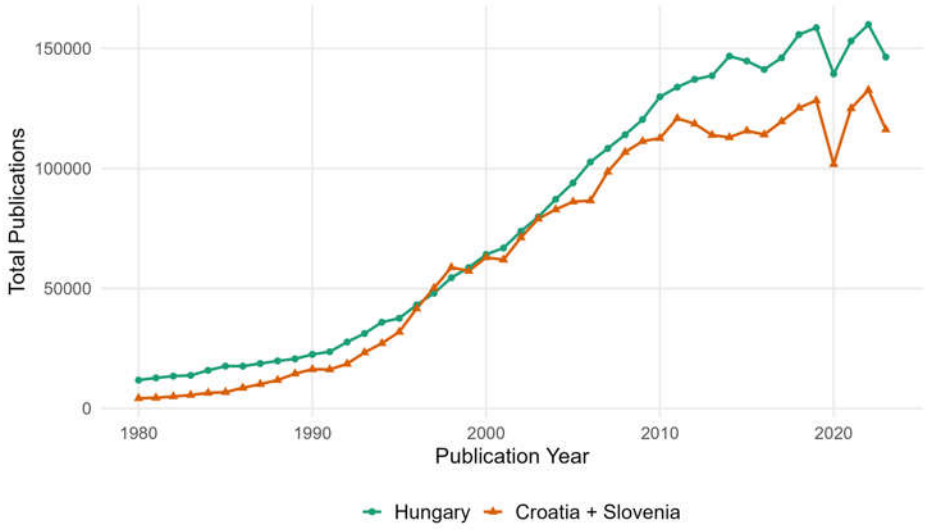
Figure 1.B.1 shows the total number of publications in each national database over 1980–2023, irrespective of any sample selection. Recorded publications follow very similar trends in Hungary and in the two control countries until 2010, with no trend break in data collection

around this date. All three series drop in 2020, showing the effect of the Covid pandemic on the overall academic production in all three countries.

Figure 1.B.2 shows the average number of publications per author and year. Hungarian authors publish more than authors in Croatia and Slovenia in the early 2000s, but the gap closes by 2008. Between 2008 and 2015, both groups have similar levels and trends. Afterwards, the average is slightly higher in the control countries. These differences in levels are absorbed by researcher fixed effects. What matters for our identification is that trends are similar just before 2011.

Figure 1.B.3 plots average publication trajectories by years since entry into academia. In all three countries, publications increase during the first 10 to 15 years of the career and decline afterwards (panel a). Conditional on publishing, output rises with seniority and stabilises after about 20 years (panel b). Levels differ across countries, with Croatian authors publishing the least, but the shape of academic productivity over the year is similar. Since cohort by year fixed effects are common to the three countries, this similarity supports our specification, and the differences in levels are again absorbed by researcher fixed effects.

Figure 1.B.1 : Total Publications per Year in National Databases

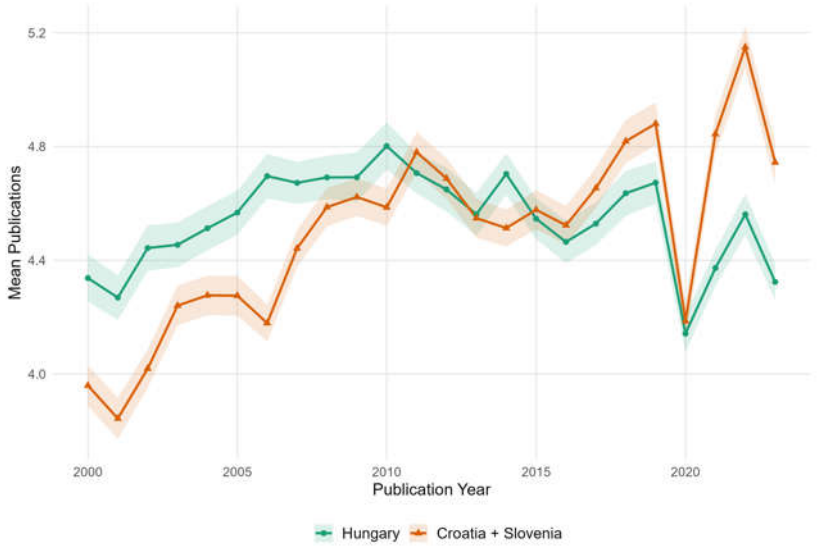


Notes: The graph plots the total number of publications recorded per year in the Current Research Information Systems (CRIS) used in the cross-country author-level analysis: MTMT (Hungary), SICRIS/COBISS (Slovenia) and CROSBI/CRORIS (Croatia). All recorded publications are included, without any sample restriction.

1.C — Academics' Political Alignment

We detail in this section how we build our samples of researchers who are likely to be perceived by the government as political opponents or as political allies. We consider three groups of perceived political opponents: researchers affiliated with the Central European University (CEU), academics linked to NGOs and activist networks targeted by the policies of Viktor Orbán's government, and researchers of the Institute of Political History (*Politikatörténeti Intézet*). We consider three groups of perceived political allies: members of the Batthyány Society of Professors, researchers affiliated with the Mathias Corvinus Collegium (MCC), and researchers who are members of Fidesz. Because these groups are small, we match their members by hand to their identifiers in the Hungarian CRIS (MTMT) and in OpenAlex, based on their publication records. To supplement and verify the data, we use researchers' personal websites, online CVs and LinkedIn profiles to collect key information, including nationality, year of entry

Figure 1.B.2 : Average Publications per Year - Full Sample



Notes: The graph shows the average number of publications per author per year in the CRIS used in the cross-country author-level analysis (Hungary, Slovenia and Croatia).

and final year of affiliation with the relevant institution. This process also allows us to exclude deceased individuals and to correct OpenAlex data when it reports wrong affiliations. For each group, we only consider the information pre-2010. This ensures that we do not inadvertently include researchers who joined the different groups in reaction to the political opposition or support to the current government.

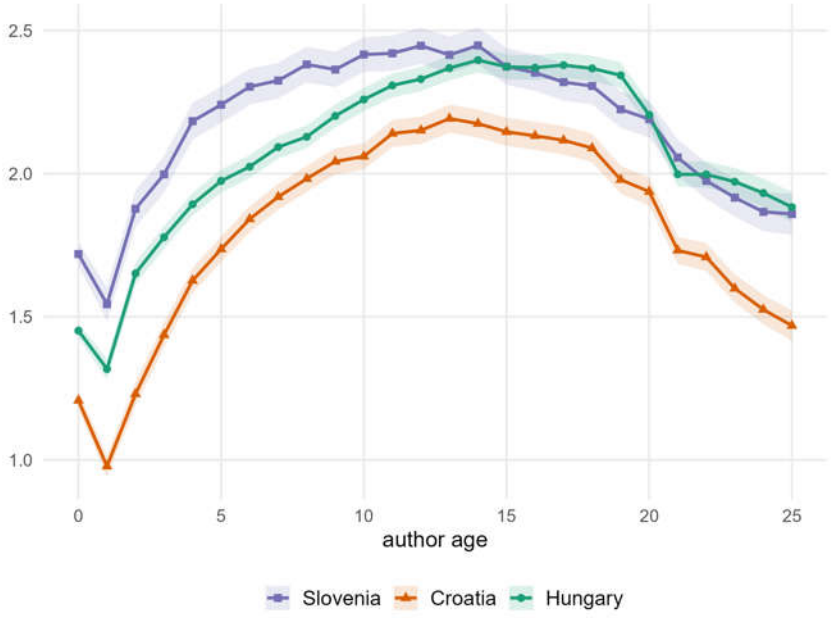
1. Perceived Political Opponents

Central European University. We extract from OpenAlex all researchers affiliated with the Central European University since its creation, a private university founded by George Soros which was forced to move most of its activities from Budapest to Vienna after the 2017 amendment to the higher education law (*Lex CEU*). We match authors to MTMT and OpenAlex by hand. For regression analyses using this sample, we restrict attention to researchers who began their affiliation with the Central European University before 2010.

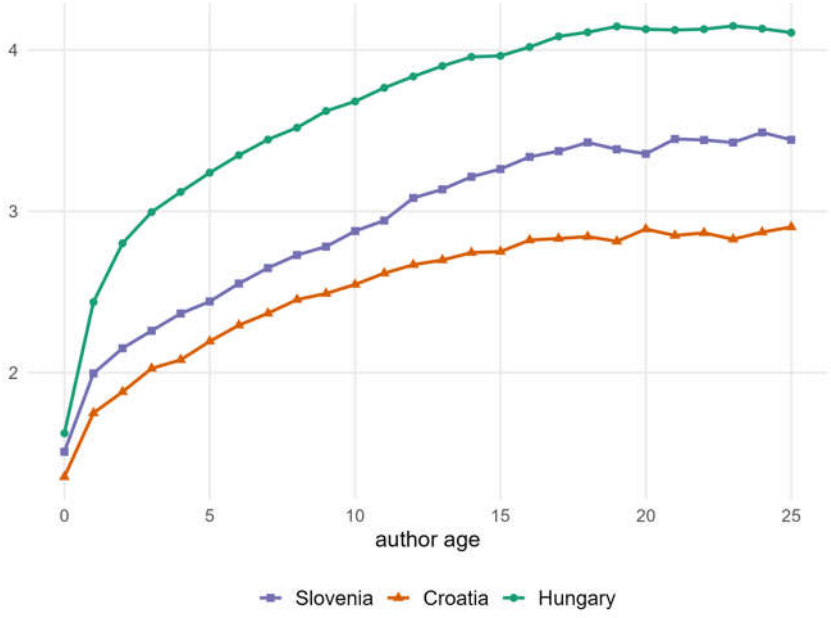
NGOs and Activists. To identify researchers linked to NGOs, we identify several NGOs that always have been critical of the government, namely the Hungarian Helsinki Committee, Amnesty International, Transparency International Hungary, the Open Society Foundations and *Menedék* (Hungarian Association for Migrants). We confirm this list of NGOs using the so-called "Soros list", a document published in 2018 by *Figyelő*, a pro-government newspaper. The list names around 200 academics and human-rights activists accused of being "mercenaries of Soros". Approximately 80 individuals from this list are indexed in the Hungarian CRIS database.

Institute of Political History. The Institute of Political History (*Politikatörténeti Intézet*) is a left-leaning research institute that is not directly funded by the state. In 2013, the Fidesz government nationalised part of its archival holdings without compensation. In 2017, the state also sued to evict the Institute from the building that it and its predecessor had occupied since the postwar period. After the Institute initially won, an extraordinary review by the *Kúria*, Hungary's supreme court, opened the way for lower courts to order its eviction in June 2020.

Figure 1.B.3 : Average Publication Trajectories per Author



(a) Average Number of Publications per Year



(b) Average Number of Publications per Year, Conditional on Publishing

Notes: The graphs plot the average number of publications per author in the CRIS used in the cross-country author-level analysis (Hungary, Slovenia and Croatia). In both graphs, the x-axis shows the number of years since entry into academia. Panel (b) only includes author-years with at least one publication.

2. Perceived Political Allies

Batthyány Society. The Batthyány Society of Professors is a conservative association of academics close to Fidesz. The full list of its members can be found on its website pbk.info.hu. As for the CEU, we match them by hand to their identifiers in MTMT and OpenAlex based on publication records, and collect the same information.

Mathias Corvinus Collegium. Founded in 1996, the Mathias Corvinus Collegium (MCC) is a private think-tank, educational institution and research centre "devoted to the flourishing of the Hungarian nation", according to the organisation. The institution is closely tied to Viktor Orbán's government. We identify researchers affiliated with MCC using the website of the institution, mcc.hu, and match them to MTMT and OpenAlex by hand.

Fidesz Members. Finally, we include researchers who were members of Fidesz, Viktor Orbán's party, before 2010. As there is no single source for this group, we compile it progressively, adding researchers whom we identify as members of the party before 2010 in the course of our research on the political repression of academics in Hungary. The list is therefore unlikely to be exhaustive. We then match these researchers to MTMT and OpenAlex data by hand.

3. Networks of Opponents, Networks of Allies

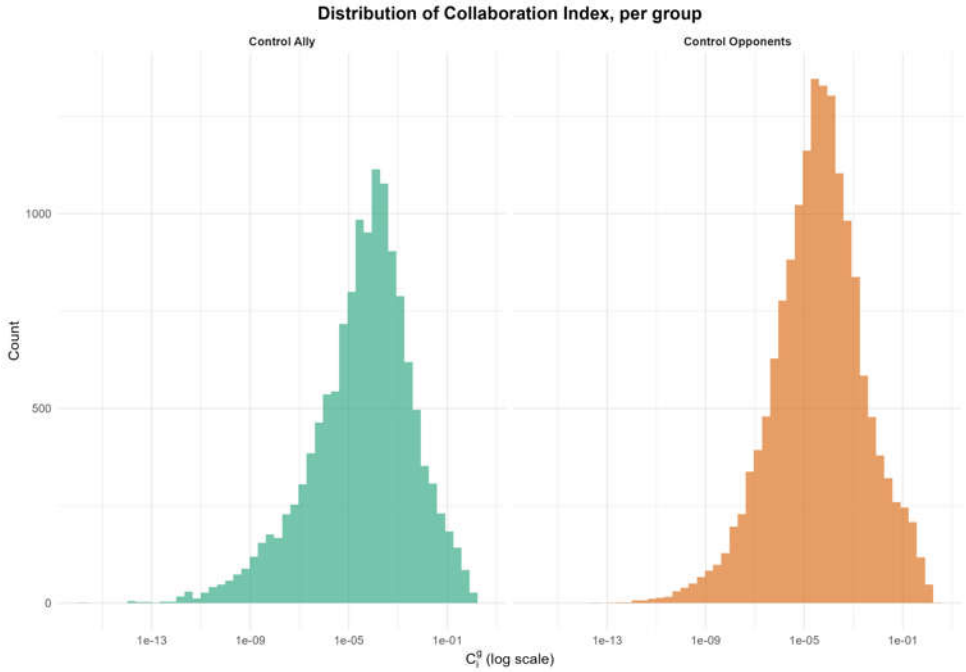
a) Network Description

Table 1.C.1 : Pre-Treatment Descriptive Statistics by Group

		Network pol. opp.		Network ally	
		Treat	Control	Treat	Control
Panel A: Non-matched sample					
Entry year	Mean	1990.64	1995.66	1993.81	1994.91
	SD	12.21	11.94	12.60	11.61
Life Sciences	Share	0.23	0.27	0.29	0.26
Social Sciences	Share	0.43	0.31	0.28	0.30
Physical Sciences	Share	0.30	0.43	0.46	0.37
Health Sciences	Share	0.28	0.23	0.21	0.33
Publications	Mean	4.89	2.87	3.98	3.43
	Median	3.00	2.00	2.00	2.00
	SD	6.60	4.07	5.86	5.10
Publications (English)	Mean	2.50	1.55	2.35	1.79
	Median	1.00	1.00	1.00	1.00
	SD	4.27	2.88	4.21	3.15
Publications (Hungarian)	Mean	2.80	1.60	2.00	1.95
	Median	1.00	0.00	0.00	1.00
	SD	4.77	3.09	4.24	3.92
Citations	Mean	59.28	34.14	60.96	34.79
	Median	7.00	1.00	4.00	2.00
	SD	211.17	216.89	303.16	116.43
N authors		5,845	4,464	4,529	1,790
Panel B: Matched sample					
Entry year	Mean	1991.81	1993.57	1994.45	1993.38
	SD	10.44	10.37	10.00	10.40
Life Sciences	Share	0.19	0.23	0.27	0.21
Social Sciences	Share	0.42	0.32	0.27	0.26
Physical Sciences	Share	0.21	0.36	0.33	0.31
Health Sciences	Share	0.26	0.19	0.19	0.30
Publications	Mean	4.59	2.90	3.57	3.37
	Median	3.00	2.00	2.00	2.00
	SD	6.22	4.06	5.14	5.06
Publications (English)	Mean	2.20	1.49	2.05	1.71
	Median	1.00	0.00	1.00	1.00
	SD	3.91	2.78	4.02	3.13
Publications (Hungarian)	Mean	2.72	1.65	1.82	1.96
	Median	1.00	0.00	0.00	0.00
	SD	4.60	3.18	3.65	4.01
Citations	Mean	54.79	36.48	63.94	35.33
	Median	6.00	1.00	4.00	2.00
	SD	216.61	239.25	326.17	113.89
N authors		3,592	2,544	1,899	1,028

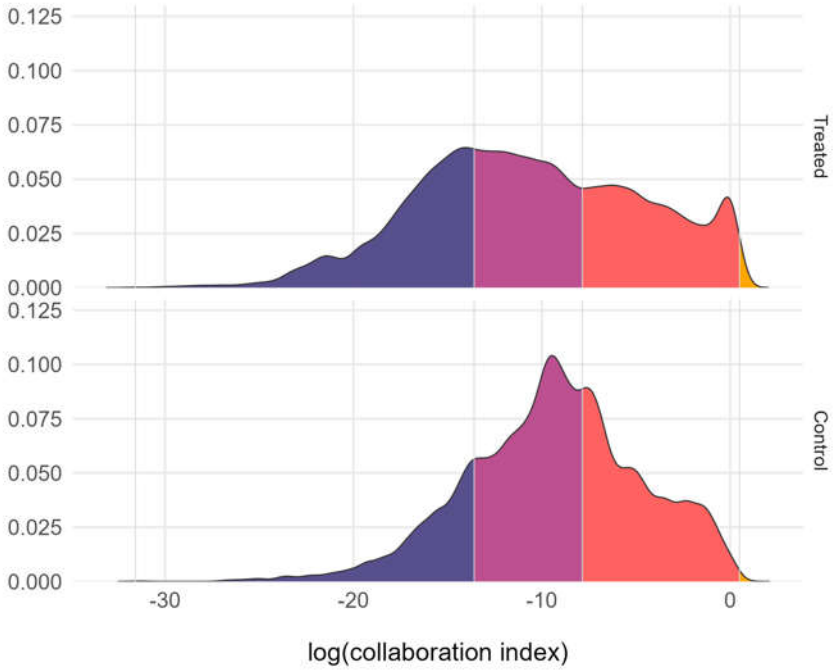
Note:
Statistics computed on the pre-treatment period (2000–2009). Entry year and institution shares are computed at the author level; publication and citation statistics over author-year observations. Panel B restricts each comparison to its exact-matched sample (matched on entry year and field).

Figure 1.C.1 : Distribution of Network Index - Control Groups



Notes: Each pane of this graph is showing the distribution of $C_t^{g,\text{net}}$ for the control groups ally researchers and the control group of opponent researchers.

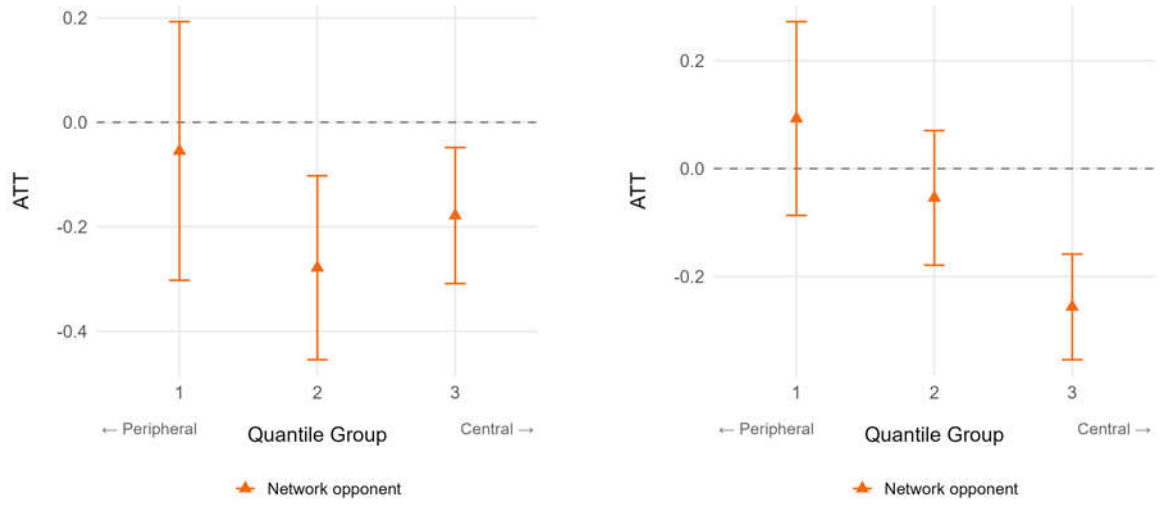
Figure 1.C.2 : Distribution of Network Index - Quantiles



Notes: This graph show the log distribution of the $C_t^{g,\text{net}}$ for the opponent network group (top pane) and the control group associated (bottom pane). Colours show the quantiles that we consider for both group in terms centrality in the network of Opponents.

b) **Heterogeneity Results**

Figure 1.C.3 : Network Proximity – Effect on Influence and Collaborations



(a) Citation count

(b) New Collaborations

Note: Results are from Fixed-Effect PPML regressions from CRIS data. The unit of observation is author by year and the sample period is 2005–2023. Treated authors are those whose collaboration index with researchers perceived as political opponents is above median. We then split them in three groups by tercile of collaboration index. Panel (a) reports the effect on citations within Hungarian CRIS. Panel (b) reports the effect on new collaborations. Control variables account for trends field and cohort dynamics over time. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

4. Targeted Research Topics

a) Topic Definition

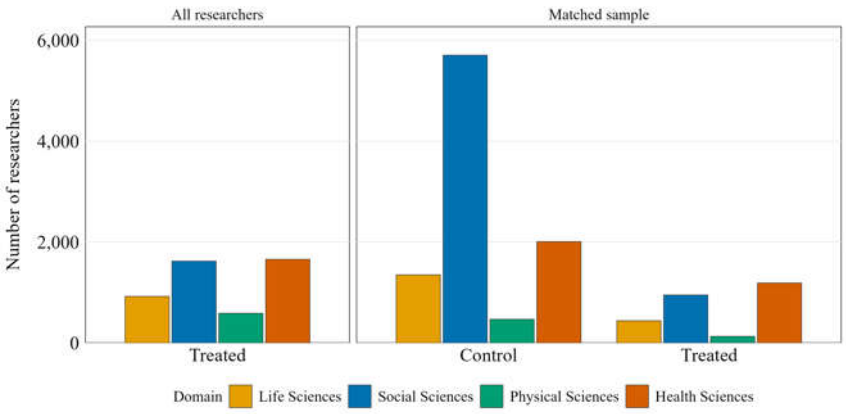
This section details how we identify publications belonging to four thematic categories: gender-related topics, rule of law and democracy, broader sensitive political topics, and climate change. Our classification relies on the presence of specific keywords and expressions in publication titles. For each category, we construct a curated list of terms commonly associated with the corresponding literature. To reduce noise, we classify an author as contributing to a given topic based on whether they published a number of these types of publications greater than three.

b) Gender

We identify gender-related publications using a set of keywords capturing the main dimensions of gender studies, including social theory, health, sexuality, and family structures. The list is designed to cover both conceptual and empirical work in the field. This topic includes several interconnected subtopics. First, gender relations and feminist theory (*gender*, *women*, *sexism*, *patriarchal*, *feminism*, *stereotype*, *toxic masculinity*, *intersectionality*) explores power structures, social norms, and inequalities shaping gendered experiences. Second, reproductive health and biomedical dimensions (*reproductive health*, *abortion*, *contraception*, *fertility*, *pregnancy*, *childbirth*, *gynecologic*, *menopause*, *uterus*, *ovary*) focuses on the biological, medical, and embodied aspects of reproduction. Third, sexuality, rights, and gender-based violence (*sexuality*, *sexual empowerment*, *sexual harassment*, *sexual violence*, *female genital mutilation*, *forced sterilization*, *partner violence*, *femicide*) addresses bodily autonomy, rights, and forms of gendered harm. Fourth, LGBTQ+ identities and gender diversity (*LGBT*, *lesbian*, *gay*, *transgender*, *non-binary*, *queer*, *intersex*, *sexual orientation*, *trans rights*) highlights issues of identity, recognition, and inclusion. Finally, family, labor, and social organization (*domestic labor*, *care economy*, *motherhood*, *maternity*, *family policy*, *child marriage*, *gendered division of labor*) examines how gender structures everyday life, institutions, and economic relations.

Figure 1.C.4 presents the distribution of researchers working on gender topics by main domain of publication. Gender topics are not confined to social sciences and humanities: although about a third of the sample publishes in such sciences, the largest group of treated researchers work in health sciences. Some also study life sciences, with a few working in physical sciences (for instance, at the intersection of chemistry and pharmaceutics).

Figure 1.C.4 : Distribution by domain of researchers working in gender topics in entire dataset, control group and treated group of the matched sample.



Notes: The graph above shows the number of authors working on gender-related topics in Hungary by domain in the entire dataset, and the number of control researchers and researchers working on gender-related topic by domain after matching. An author who publishes in two different domains is counted once per domain. We consider an author to be working on gender topics if she has at least three publications on gender-related topics between 2000 and 2005. Gender publications are defined using the method detailed in Appendix 1.C.4.

c) Rule of Law

To capture publications related to the rule of law and democratic institutions, we rely on a set of keywords commonly used in discussions of institutional quality, civil liberties, and authoritarian dynamics. These terms are frequently mobilised in academic work critically engaging with illiberal policies and democratic backsliding. This topic includes several interconnected subtopics. First, fundamental rights and legal principles (*human rights*, *civil rights*, *fundamental rights*, *rule of law*, *constitutional law*, *legal equality*, *fair trial*, *access to justice*) refers to the normative and institutional foundations that protect individual freedoms and ensure legal accountability. Second, political ideologies and authoritarian dynamics (*authoritarian*, *nationalism*, *political conservatism*, *populism*, *radical right*, *far right*, *autocracy*, *illiberalism*, *fascist*) captures ideological currents and regimes that challenge or reshape democratic norms. Third, democratic erosion and institutional capture (*democratic backsliding*, *erosion of democracy*, *judicial capture*, *politicised judiciary*, *constitutional crisis*, *manipulated elections*, *electoral fraud*, *state capture*) highlights processes through which democratic institutions are weakened or subverted. Fourth, surveillance, repression, and state control (*state surveillance*, *mass surveillance*, *political repression*, *police state*, *censorship*, *propaganda*, *arbitrary detention*, *state violence*, *enforced disappearance*) focuses on coercive tools used to monitor, control, and silence populations. Fifth, corruption and governance practices (*corruption*, *kleptocra*, *cronyism*, *clientelism*, *nepotism*, *bribery*, *patronage system*, *abuse of power*) examines informal and illegal mechanisms of power consolidation. Finally, freedoms, civil society, and academic autonomy (*freedom of speech*, *freedom of expression*, *freedom of the press*, *freedom of assembly*, *academic freedom*, *university autonomy*, *media freedom*, *shrinking civic space*, *chilling effect*) addresses the conditions necessary for open debate, knowledge production, and democratic participation.

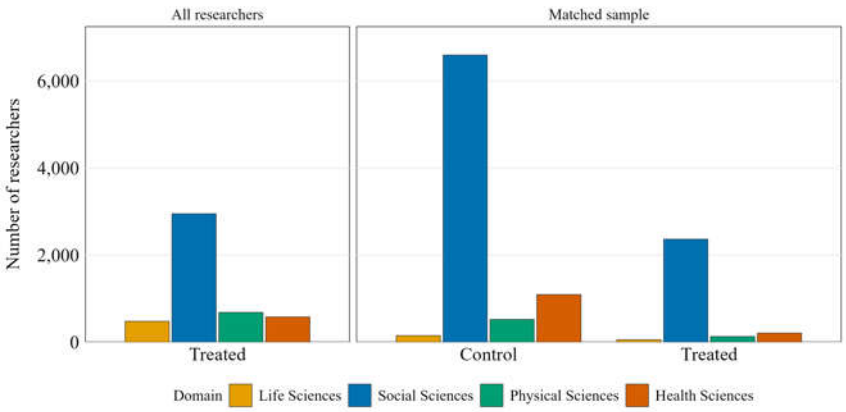
d) Sensitive Topics

We define a broader category of sensitive political topics encompassing social justice, political conflict, and authoritarian practices. The associated keywords capture themes that are often politically salient, contested, or ideologically charged. This topic includes several interconnected subtopics. First, migration, borders, and global inequalities (*migrant*, *refugee*, *asylum*,

borders, *internationalism*) highlights transnational movements of people and the political and ethical issues surrounding displacement and border regimes. Second, economic systems and class struggle (*capitalism*, *neoliberal*, *exploitation*, *social inequality*, *economic justice*, *income redistribution*, *anticolonialism*, *anti-imperialism*) focuses on critiques of economic structures and their role in producing inequality and domination. Third, labor, workers, and collective action (*workers*, *labor condition*, *trade union*, *workers' rights*, *workers' movement*, *class struggle*, *proletariat*) examines material conditions, labor organization, and social conflict. Fourth, left-wing ideologies and revolutionary thought (*marxism*, *communism*, *socialism*, *anarchism*, *left-wing*, *progressive politics*, *revolutionary theory*) captures political doctrines and movements advocating systemic transformation. Fifth, identity, discrimination, and social justice (*racism*, *antisemitism*, *islamophobia*, *sexism*, *intersectional*, *racial discrimination*, *equal opportunity*, *solidarity*) addresses inequalities based on identity and struggles for inclusion and rights. Sixth, historical violence, genocide, and memory (*holocaust*, *genoci*, *extermination*, *deportation*, *slave trade*, *denazification*) reflects on mass violence, collective trauma, and historical accountability. Seventh, authoritarianism and political repression (*autocracy*, *totalitarianism*, *dictator*, *fascism*, *police state*, *repression*, *surveillance state*, *cult of personality*) examines coercive regimes and mechanisms of control. Eighth, democratic erosion and institutional abuse (*democratic backsliding*, *judicial capture*, *electoral fraud*, *state capture*, *corruption*, *clientelism*, *abuse of power*) highlights processes undermining democratic governance. Finally, civil liberties and resistance (*freedom of speech*, *press freedom*, *censorship*, *political persecution*, *torture*, *political prisoner*, *exile*) emphasizes the restriction of freedoms and the struggles against oppression.

The group of researchers working in these fields is disproportionately working in social sciences, but not exclusively: Figure 1.C.5 shows that a relatively large number works in the other three domains of science. However, in our causal analysis, the vast majority of researchers working on sensitive topics that we are able to match to a comparable neutral researcher study social sciences.

Figure 1.C.5 : Distribution by domain of researchers working in sensitive topics in entire dataset, control group and treated group of the matched sample.



Notes: The graph above shows the number of authors working on sensitive topics in Hungary by domain in the entire dataset, and the number of control researchers and researchers working on sensitive topic by domain after matching. An author who publishes in two different domains is counted once per domain. We consider an author to be working on sensitive topics if she has at least three publications on sensitive topics between 2000 and 2009. Gender publications are defined using the method detailed in Appendix 1.C.4.d.

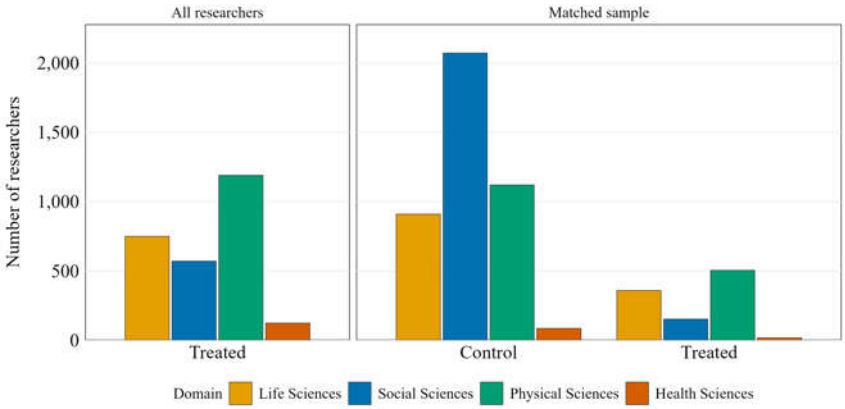
e) Climate change

Finally, we define a topic centred on climate change and environmental politics. The associated keywords capture themes that are scientifically established yet politically contested, spanning both the physical dimensions of environmental change and its social, economic, and

cultural dimensions. First, core climate concepts and discourse (*climate change*, *climate crisis*, *global warming*, *climate policy*, *climate risk*, *climate change denial*, *climate skepticism*) captures the central vocabulary of this literature, including both the framing of the problem and the contestation of climate science itself. Second, environmental science and Earth systems (*greenhouse gases*, *carbon emissions*, *carbon cycle*, *air pollution*, *sea level rise*, *permafrost*, *natural disaster*) refers to the physical processes, mechanisms, and observable consequences of environmental change. Third, biodiversity and ecosystem decline (*biodiversity*, *loss of biodiversity*, *biodiversity crisis*, *biodiversity conservation*) addresses the erosion of ecosystems and species, together with efforts at their protection. Fourth, causes and structural drivers (*fossil fuels*, *extractivism*, *overconsumption*, *industrial agriculture*, *land degradation*, *ecological debt*, *climate colonialism*) identifies the economic activities and structural dynamics held responsible for environmental degradation, including critiques linking ecological harm to global power asymmetries. Fifth, energy transition and technological solutions (*renewable energy*, *green energy*, *carbon neutrality*, *carbon capture*, *green transition*, *sustainable development*, *recycl*) covers proposed responses to the climate crisis, from technological innovation to sustainability practices and resource management. Sixth, climate justice, mobilization, and displacement (*climate justice*, *environmental justice*, *climate refugees*, *just transition*, *climate activism*, *climate strike*, *Fridays for Future*, *Extinction Rebellion*) highlights the distributional and ethical dimensions of climate change, the displacement of populations, and the social movements demanding political action. Finally, environmental philosophy, culture, and theory (*anthropocene*, *ecocide*, *eco-anxiety*, *ecofeminism*, *deep ecology*, *degrowth*, *post-growth*, *planetary boundaries*) captures conceptual and critical frameworks that rethink the relationship between humanity and nature, including radical critiques of growth-based economic models.

This group particularly allows us to study researchers who might be targeted due to their field of study in physical sciences. We plot the number of researchers working on these topics in Hungary by scientific domain in Figure 1.C.6. Both in all MTMT and in the matched sample, physical sciences are the largest group, and the second largest is life sciences.

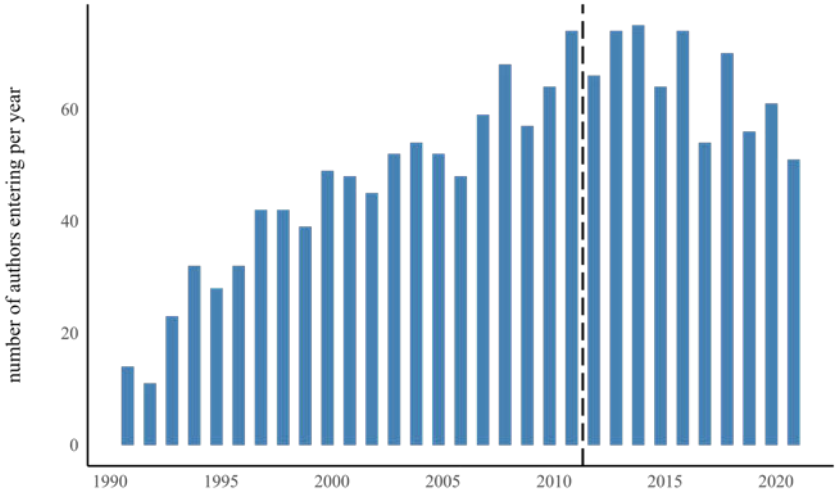
Figure 1.C.6 : Distribution by domain of researchers working in climate-related topics in entire dataset, control group and treated group of the matched sample.



Notes: The graph above shows the number of authors working on climate-related topics in Hungary by domain in the entire dataset, and the number of control researchers and researchers working on climate-related topic by domain after matching. An author who publishes in two different domains is counted once per domain. We consider an author to be working on climate-related topics if she has at least three publications on climate-related topics between 2000 and 2009. Gender publications are defined using the method detailed in Appendix 1.C.4.d.

f) **Descriptive Statistics on New Researchers**

Figure 1.C.7 : Number of entering authors working on gender topics.



Notes: The graph above shows the number of new authors working on gender-related topics between 1990 and 2020. The entering year is defined as the year of an author's first publication in our database. We use a five-year forward window to identify only the first publications of each author's career. We define new authors working on gender topics as those with more than one publication on gender-related topics within this five-year period. Gender publications are defined using the method detailed in Appendix 1.C.4.

1.D — Theoretical Framework: Exit, Voice, Loyalty or Self-censorship

The empirical analysis shows that the chilling effect of political pressure spreads beyond the researchers who are directly targeted. We develop a simple framework to account for the behaviour of differently exposed researchers: those in the network of perceived opponents, exposed to weaker pressure than the directly targeted but stronger than neutral ones, and all Hungarian researchers, subjected to rising pressure over time.

1. Setup

Consider a static model in which a conflict arises, in an organisation or a country, between the leadership and individuals. Pressure p is imposed on those who disagree, proportionally to their disagreement. Individuals all receive a wage y and are indexed along two dimensions. Perceived opposition θ_i, normalised on $[-1, 1]$ with a neutral value at 0, measures how far the values or goals of individual i differ from the leadership's. Returns to migration m_i capture the gains and costs of exiting, and can likewise be of either sign. Four mutually exclusive strategies are available. *Loyalty* leaves the individual enjoying her agreement, or suffering from her dissent, in proportion to pressure. *Self-censorship* changes the direction of her research at a fixed cost δ and, for simplicity, avoids the disutility of disagreement entirely. *Voice* costs a fixed r plus a component convex in pressure[100], and returns a utility increasing in p, in a return to dissent v (visibility, reputation among peers, moral satisfaction), and in m_i, a higher outside option conferring bargaining power. *Exit* costs F and frees her from her disagreement altogether. Individual i, characterised by a pair (m_i, θ_i), chooses whichever of the four strategies pays most:

$$U_i = \begin{cases} U_i^L & = y - p\theta_i \\ U_i^C & = y - \delta \\ U_i^V & = y - p\theta_i + p(m_i + v)\theta_i - p^2 - r \\ U_i^E & = y + m_i - F \end{cases}$$

The wage y is common to all strategies and individuals: it plays no role in the comparison and we omit it, with the index i, from now on.

Assumption 1 (Parameters)**.** *$p > 0$, $\delta > 0$, $r > 0$, $F > 0$, and v is large enough that $m + v > 0$ over the support of m. Disagreement and returns to migration are distributed on $[-1, 1]^2$.*

Assumption 2. *$r \geq \delta$ i.e., voicing is more costly than self-censoring.*

The condition $m + v > 0$ makes the bargaining power an individual mobilises when she voices non-negative, so that voicing is never a way for an aligned and immobile individual to gain utility. Assumption 2 states that, for an individual whose disagreement has been neutralised, dissent is more expensive than a change of agenda.

2. Solving the Model

The solution rests on a single observation. Fix m: each payoff is affine in θ, with slope $-p$ for loyalty, 0 for self-censorship and exit, and $p\,\kappa(m)$ for voice, where

$$\kappa(m) \equiv m + v - 1. \tag{1.D.1}$$

The quantity $\kappa(m)$ is the individual's *net bargaining power*: the return she extracts from dissent, $m + v$, less the one unit of pressure she keeps bearing by remaining exposed. If $\kappa(m) > 0$, the

[100] For instance, Alabrese, Capozza, and Garg (2024) show that researchers suffer a loss in public trust of their work when they express political opinions online.

payoff to voice rises with θ and dissent is the strategy of the most exposed and best connected; if $\kappa(m) < 0$, an exposed dissenter is worse off the more she disagrees and speaking cannot compete with falling silent.

Two consequences follow. First, self-censorship and exit are indistinguishable in θ: both are flat, so one dominates the other everywhere, and which one depends on m alone. Reorienting the agenda costs δ, leaving costs $F - m$. The cheaper of the two, which we call *silence*, costs

$$\Delta(m) \equiv \min\{\delta,\ F - m\}, \tag{1.D.2}$$

and it is self-censorship when $m < m^*$, exit when $m > m^*$, where

$$m^* \equiv F - \delta. \tag{1.D.3}$$

The threshold m^* involves neither θ nor p: it is the mobility at which the cost of shifting agenda equals the net cost of leaving, and the vertical frontier of Figure 1.D.1. Second, the comparison reduces to a three-way one between loyalty, silence — of utility $U_i^S = y - \Delta(m_i)$ — and voice. With slopes ordered $-p < 0 < p\kappa(m)$, the upper envelope is convex and piecewise linear, so the optimal strategy is monotone in disagreement: loyalty at the bottom, silence in the middle, voice at the top.

a) **The Frontiers**

Three comparisons remain, each linear in θ:

$$\text{loyalty} \geq \text{silence} \iff p\,\theta \leq \Delta(m), \tag{1.D.4}$$

$$\text{voice} \geq \text{silence} \iff p\,\kappa(m)\,\theta \geq p^2 + r - \Delta(m), \tag{1.D.5}$$

$$\text{voice} \geq \text{loyalty} \iff p\,(m+v)\,\theta \geq p^2 + r. \tag{1.D.6}$$

Each defines a threshold in disagreement,

$$\theta_1(m) \equiv \frac{\Delta(m)}{p}, \qquad \theta_2(m) \equiv \frac{p^2 + r - \Delta(m)}{p\,\kappa(m)}, \qquad \theta_3(m) \equiv \frac{p^2 + r}{p\,(m+v)}, \tag{1.D.7}$$

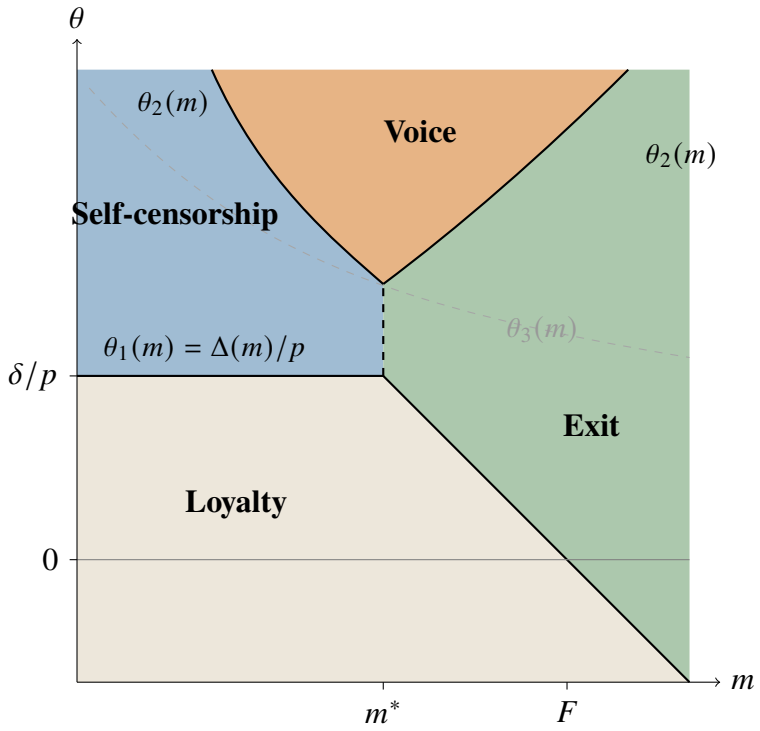
separating loyalty from silence, silence from voice, and loyalty from voice, with the convention $\theta_2(m) = +\infty$ whenever $\kappa(m) \leq 0$. The loyalty ceiling θ_1 is flat at δ/p for the immobile and falls at rate $1/p$ once m exceeds m^*, the two branches meeting there since $F - m^* = \delta$. The voice frontier moves the other way, falling below m^* and rising above it: it is V-shaped with a minimum above the kink in θ_1, and the voice region is a wedge opening at m^*.

b) **Partition of the (m, θ) Plane**

Proposition 1 (Optimal strategies)**.** *Under Assumption 1 and Assumption 2, and up to the frontiers themselves, an individual* (m, θ)

- *voices if* $\kappa(m) > 0$ *and* $\theta \geq \max\{\theta_2(m), \theta_3(m)\}$*;*
- *stays loyal if* $\theta \leq \min\{\theta_1(m), \theta_3(m)\}$*;*
- *is silent otherwise — self-censoring if* $m < m^*$*, exiting if* $m > m^*$*.*

Figure 1.D.1 : Optimal Strategy in an "Exit, Voice, Loyalty or Self-censorship" Game



Notes: Optimal strategy as a function of perceived dissent θ and returns to migration m. The figure is schematic: the axes carry no scale, only the ordering of the frontiers being meaningful. When $\delta\,\kappa(m^*) \leq p^2 + r - \delta$ the loyalty and voice regions do not touch, as drawn here; at lower pressure the self-censorship band closes and dissenters come directly from the ranks of the loyal, along θ_3. Below $\theta = 0$ the descending branch of θ_1 is the pure-migration margin $m = F - p\theta$.

Voicing therefore requires $m > 1 - v$: an individual whose bargaining power is below one never voices.[101]

For $\theta \leq 0$ only loyalty and exit are chosen, self-censorship being pointless and voice lying below loyalty by (1.D.6). Aligned researchers leave for economic reasons only, and their exit frontier $m \geq F - p\theta$ moves *outwards* as pressure rises: pressure retains the aligned at the margin.[102]

c) Voice Confined to Intermediate Pressure

The pivotal mobility $m^* = F - \delta$ is free of p: pressure alone never pushes an individual from self-censorship into exit, or the reverse. The loyalty ceiling falls, since $\Delta(m)$ does not depend on p and $\theta_1 = \Delta(m)/p$ is strictly decreasing: the loyalty region shrinks monotonically, towards self-censorship for $m < m^*$ and towards exit for $m > m^*$. This is the first mechanism

[101] Fix m. If $\kappa(m) \leq 0$, $U^V - U^S = \kappa(m)\theta - (p^2 + r - \Delta(m))$. The first term is non-positive for any $\theta \geq 0$, and $\Delta(m) \leq \delta \leq r$ by Assumption 2, so the term in parentheses is always positive: $U^V \leq U^S$, and an individual with non-positive $\kappa(m)$ always prefers silence to voicing. She prefers voicing to loyalty if $\theta \geq \theta_3$, but then prefers silence to loyalty as well, since $\kappa(m) \leq 0$ means $m + v \leq 1$, so that $\theta_3 - \theta_1 = \frac{p^2 + r - \Delta(m)(m+v)}{p(m+v)} \geq 0$ by the same Assumption 2; and (1.D.6) ensures loyalty is preferred to voice for a negative θ. The only binding threshold for $\kappa(m) \leq 0$ is thus θ_1: loyalty below, silence above. For $\kappa(m) \geq 0$ the ordering of the thresholds decides. The individual remains loyal if θ falls below the lowest of θ_1 and θ_3, the minimum in the statement. If $\theta_1 \geq \theta_2$, silence is dominated by loyalty, θ_2 cannot bind, and she voices if $\theta \geq \theta_3$ and remains loyal otherwise. If $\theta_2 > \theta_1$, loyalty is dominated by silence, and she is silent if $\theta \leq \theta_3$ and voices otherwise.

[102] We show empirically that exit rises with pressure for neutral researchers too, whereas here a researcher at $\theta = 0$ leaves whenever $m \geq F$, at any level of pressure. The gap is partly measurement, since θ is *perceived* opposition and few researchers sit exactly at zero, and partly the absence, in this design, of any deterioration of the research environment for those who stay.

behind both the spread of self-censorship and the political component of exit, the direction in which an individual leaves loyalty being fixed by her mobility relative to m^*.

The two voice frontiers move in opposite directions over different ranges of pressure:

$$\theta_2(m) = \frac{1}{\kappa(m)}\left(p + \frac{r - \Delta(m)}{p}\right), \qquad \theta_3(m) = \frac{1}{m+v}\left(p + \frac{r}{p}\right),$$

their derivatives in p have the signs of $p^2 - r + \Delta(m)$ and $p^2 - r$, so both are convex and U-shaped in pressure, with minima at $\sqrt{r - \Delta(m)}$ and $\sqrt{r}$. The binding frontier is $\max\{\theta_2, \theta_3\}$: θ_3 at low pressure and θ_2 at high pressure, the switch occurring at

$$p_s(m) = \sqrt{(m+v)\,\Delta(m) - r}\,. \tag{1.D.8}$$

Being the maximum of two convex functions, it is itself convex, and diverges at both ends: $\theta_3 \to +\infty$ as $p \to 0$, $\theta_2 \to +\infty$ as $p \to \infty$. The set of pressures at which the voice region is non-empty is therefore an interval, and voicing is an intermediate strategy. What changes with pressure is the strategy against which voicing competes. At low pressure silence is not worth its cost and the comparison is with loyalty: θ_3 binds and falls with p, pressure making loyalty painful faster than it makes dissent dangerous, so the voice region expands. At high pressure the comparison is with silence: θ_2 binds and rises, and the region contracts until it vanishes at $\bar{p}$. Dissent is thus confined to an intermediate band, at its widest around $\sqrt{r}$.[103]

As $p \to 0$ the model collapses to pure migration, exit being decided by $m \geq F$ alone; as $p \to \infty$, $\theta_1 \to 0$ and it becomes a partition between the aligned, who stay, and the dissenting, silent in one of the two available ways. A period of autocratisation should therefore display a hump in dissent rather than a monotone decline.

d) Loyalty first, Voice, then Silence

The previous subsection describes how the frontiers move, not the order in which a given individual changes strategy. Fix (θ, m) with $\theta > 0$ and $m < F$ and write each deviation from loyalty as a function of pressure:

$$g^S(p) = p\theta - \Delta(m), \qquad g^V(p) = p\theta(m+v) - p^2 - r. \tag{1.D.9}$$

The individual plays whichever deviation is larger, and stays loyal while both are negative. The gain to silence is an increasing line, negative for a positive θ only if $\Delta(m) > 0$ and hence $m < F$: an individual whose outside option already exceeds the cost of leaving departs at any pressure and has no loyal phase. $\Delta(m)$ does not depend on p: pressure decides *when* an individual falls silent, never *in which way*, that choice being driven by (1.D.2) alone.

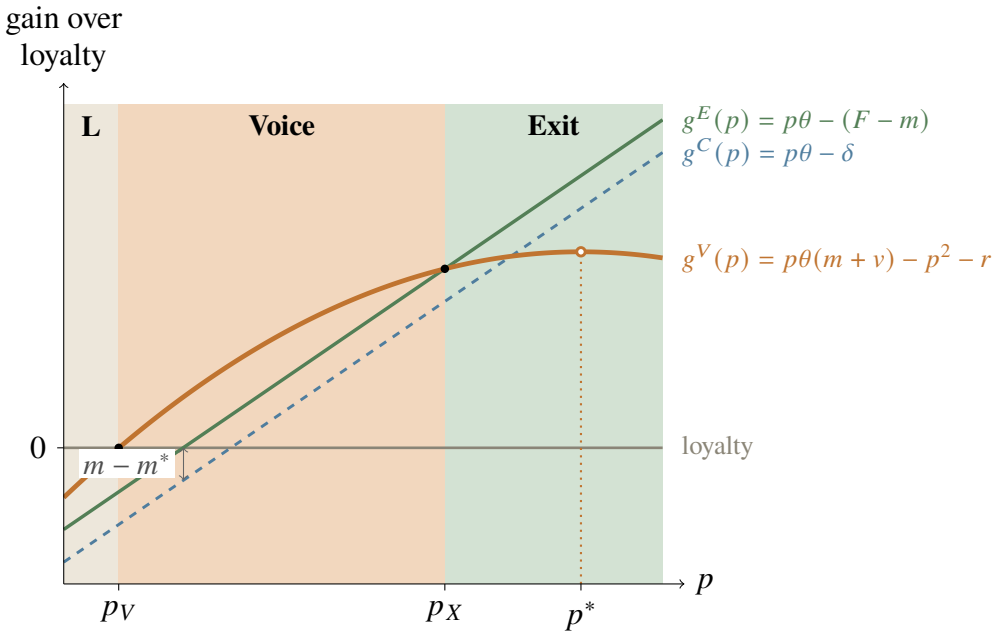
The gain to voice is concave. It starts at $g^V(0) = -r$, peaks at

$$p^* = \frac{\theta(m+v)}{2}, \tag{1.D.10}$$

where the marginal return to bargaining power $\theta(m+v)$ equals the marginal cost of dissent $2p$, and declines thereafter. It is positive only between the two roots of $d^v(p) = p^2 - p\theta(m+v) + r = 0$, real if and only if $\theta(m+v) \geq 2\sqrt{r}$: bargaining power must clear a threshold set by the fixed

[103] Since θ_2 is smallest at m^* and disagreement is bounded above by one, the voice region is non-empty only if $\theta_2(m^*) \leq 1$, that is if $p^2 - \kappa(m^*)\,p + r - \delta \leq 0$. Under Assumption 2 the constant term $r - \delta$ is non-negative, so this quadratic has two non-negative roots whenever they are real and the condition defines a bounded interval $[\underline{p}, \bar{p}]$: the voice region is empty at low as well as at high pressure, and at every pressure when the fixed cost of speaking is too large relative to the bargaining power available at m^*.

Figure 1.D.2 : Timing of Responses



Notes: Optimal responses as a function of political pressure p. The axes carry no scale, only the ordering of the frontiers being meaningful. The vertical distance between the two silent gains is $m - m^*$: the individual drawn here, for whom g^E lies above g^C, is mobile enough to leave rather than reorient.

cost of speaking before dissent is worth its price. That fixed cost also places the curve below loyalty at zero pressure; without it there would be no loyal phase at all. A concave curve and an increasing line cross at most twice, so the order in which the deviations take the lead is fixed: loyalty, then voice, then silence. The individual begins to voice at p_V, the smaller root of $d^v(p)$, and falls silent at p_X, the larger, self-censoring if $m < m^*$ and exiting if $m > m^*$.[104]

Figure 1.D.2 draws the gains for a mobile researcher. Voicing is given up not because it has stopped paying but because exit has started paying more, and the two have different empirical signatures: a fall in the intensity of dissent against a rise in departures among those who were dissenting. Read as an autocratisation episode in which pressure rises over a sequence of mandates, this predicts dissent early and exit later, with a lag proportional to $p_X - p_V$.[105]

3. Predictions

The framework delivers six predictions. The first four follow formally from the model; the last two combine them with an assumption on the mapping from network position to perceived opposition.

1. *Sorting with disagreement.* For a given outside option, the strategy is monotone in disagreement: low θ_i stay loyal or leave for purely economic reasons; at higher θ_i, the immobile self-censor and the sufficiently mobile voice or exit.
2. *Self-censorship rises with pressure.* As p increases, the threshold above which loyalty becomes too costly falls, so individuals with an ever smaller θ_i change agenda. At low pressure the effect is ambiguous, the ceiling of the band falling with its floor as

[104] This requires that silence still be dominated when voice becomes profitable, $\theta p_V < \Delta(m)$, and already profitable when voice is overtaken, $\theta p_X \geq \Delta(m)$; both hold in Figure 1.D.2. When the second fails, the voice window closes before silence pays and a fourth phase appears: the individual voices, returns to loyalty, and falls silent only at a much higher pressure.

[105] The model being static, the sequence is a comparative-statics path and presumes a constant (θ, m); and since p_V and p_X depend on (θ, m), the aggregate profile is a mixture over the population rather than a scaled-up Figure 1.D.2.

voicing expands; self-censorship grows unambiguously only past a pressure of the order of $\sqrt{r}$.

3. *Voicing is an intermediate strategy.* Voicing is attractive only when pressure is high enough to make silence costly, and its returns fall beyond a given value. Individuals start voicing and then fall back on self-censorship or exit, depending on their mobility — across individuals at a given pressure, and along the pressure path of one individual: dissent early, exit later.
4. *Exit has a political component.* The loyalty–exit frontier depends on p: exit increases with pressure, including among those who would never voice.
5. *Attenuation with network distance.* Since θ decays from the core of the opponent network outward, every effect fades with distance from the core.
6. *Isolation is a composition effect.* The mobile leave, the immobile retreat, and the aligned become a growing share of those who remain: no actor needs to set out to isolate the system, the outcome being the sum of individually rational strategies. If m is related to quality of work, the composition effect lowers the average quality of the stayers.

Our findings are in line with this framework: exit becomes more likely as pressure increases for all categories of researchers, including the neutral ones, with no significant effect during the first mandate, while self-censorship rises over time. Self-censorship is thus not confined to those at the margins, and may be as frequent a strategy as migration — a pattern that could extend beyond Hungarian researchers to illiberal regimes at large.

4. Population Shares

Section 1.D.2.d follows one individual, but the objects we measure are counts. Aggregating the partition over the population turns the frontiers into shares, and delivers one prediction the individual analysis does not. Let (m, θ) be uniform on $[-1, 1]^2$, so that the density is $1/4$, and write

$$\lambda(m) \equiv \min\{\theta_1(m), \theta_3(m)\}, \qquad \upsilon(m) \equiv \max\{\theta_2(m), \theta_3(m)\} \tag{1.D.11}$$

for the loyalty ceiling and the voice floor, the min and the max covering both configurations of Proposition 1. Each share is a quarter of the integrated vertical extent of its region: loyalty below λ, voice above υ, silence in between, split at m^*. Under $p > \delta$, $p > 1 - F$ and $p^2 + r < F + v - 1$ — which place the loyalty ceiling inside the support and give θ_2 its V-shape — the four integrals are elementary. With $B \equiv (p^2 + r - \delta)/p$, $C \equiv p^2 + r - F - v + 1$, $\kappa^* \equiv \kappa(m^*)$, and $[a, b]$ the interval of mobility over which voicing occurs, $a = \max\{-1, B - v + 1\}$ and $b = \min\{1, (F - p^2 - r + p(v - 1))/(1 - p)\}$,

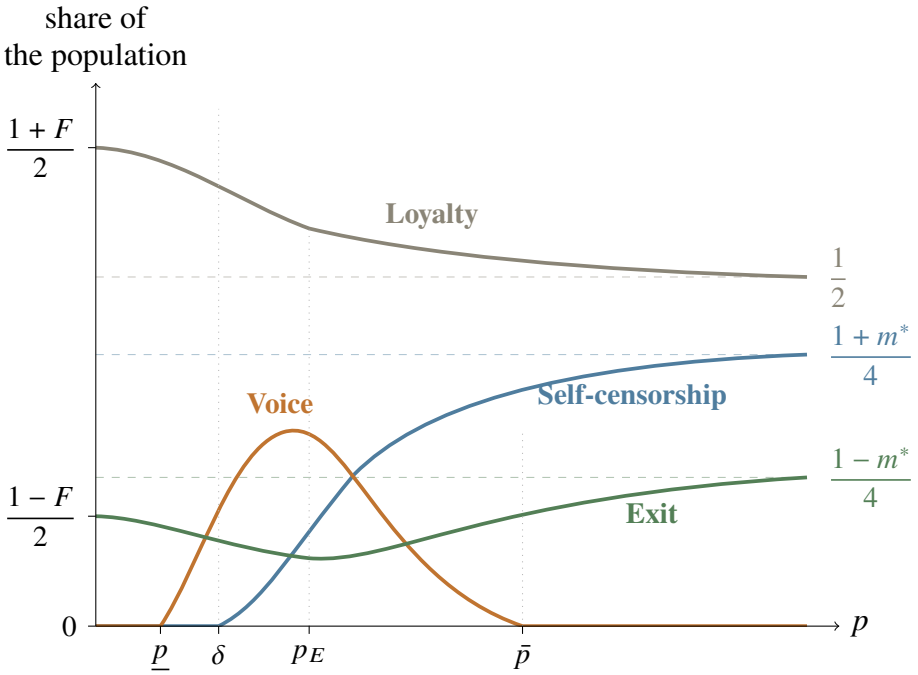
$$S_L = \frac{1}{2} + \frac{1}{4p}\left[\delta(1 + F) - \frac{\delta^2 + (1 - F)^2}{2}\right] \tag{1.D.12}$$

$$S_C = \frac{1}{4}\left[(1 + a) + B \ln \frac{\kappa^*}{\kappa(a)} - \frac{\delta(1 + m^*)}{p}\right], \tag{1.D.13}$$

$$S_V = \frac{1}{4}\left[(b - a) - B \ln \frac{\kappa^*}{\kappa(a)} - \frac{b - m^*}{p} - \frac{C}{p} \ln \frac{\kappa(b)}{\kappa^*}\right] \tag{1.D.14}$$

$$S_E = \frac{1}{4}\left[(1 - b) + \frac{b - m^*}{p} + \frac{C}{p} \ln \frac{\kappa(b)}{\kappa^*} - \frac{\delta^2 - (1 - F)^2}{2p}\right]. \tag{1.D.15}$$

Figure 1.D.3 : Population Shares as a Function of Political Pressure



Notes: Share of the population playing each strategy as a function of political pressure p, for (m, θ) uniform on $[-1, 1]^2$. The axes carry no scale, only the ordering of the curves and of the turning points being meaningful; the limits are exact. Loyalty falls hyperbolically from $(1 + F)/2$ to $1/2$, by (1.D.13). Self-censorship is nil until $p = \delta$, below which the ceiling δ/p lies above the support of θ, and rises to $(1 + m^*)/4$. Voice is confined to $[\underline{p}, \bar{p}]$ and hump-shaped, by Proposition 2. Exit begins at its pure-migration level $(1 - F)/2$, falls to a minimum at p_E and rises to $(1 - m^*)/4$, ending above its initial level iff $F + \delta > 1$. The drawing takes $m^* > 0$.

They sum to one, every logarithm cancelling in pairs.[106] Loyalty is exactly hyperbolic in pressure, $S_L = \frac{1}{2} + K_L/p$, with K_L a function of δ and F alone: neither r nor v enters, the loyalty ceiling being set by the cost of silence and never by the voice frontier.

The limits are read off the square. As $p \to 0$ the ceiling $\Delta(m)/p$ diverges and only mobility decides: $S_L \to (1 + F)/2$, $S_E \to (1 - F)/2$. As $p \to \infty$ the ceiling collapses and loyalty is the half-square $\theta \leq 0$, so $S_L \to 1/2$: the aligned never leave, loyalty paying them an unbounded $p|\theta|$ against an exit payoff bounded by $1 - F$. The dissenting half is then split by m^* alone, $S_C \to (1 + m^*)/4$ and $S_E \to (1 - m^*)/4$. Voice vanishes at both ends, and this is general.

Proposition 2 (Aggregate hump in dissent)**.** *For any distribution of (m, θ) with full support, $S_V(p) \to 0$ as $p \to 0$ and as $p \to \infty$, so the aggregate profile of dissent is hump-shaped whenever the voice region is non-empty at some pressure.*[107]

This closes the gap left by Section 1.D.2.d, where the hump is established for one individual only.

Figure 1.D.3 plots the four shares. Loyalty declines and self-censorship rises throughout, and dissent traces the hump. Exit does neither: it falls over the first stretch of the range and only then climbs. Two forces produce the dip: the exit margin for the aligned, $m \geq F - p\theta$,

[106] The antiderivatives are $\int \theta_1 = \delta m/p$ and $[Fm - m^2/2]/p$ on the two branches, $\int \theta_2 = B \ln \kappa(m)$ and $[m + C \ln \kappa(m)]/p$, using $\theta_2 = [1 + C/\kappa(m)]/p$ to the right of m^*; truncation is where $\theta_2 = 1$. Where $\delta(m^* + v) > p^2 + r$ the silence band closes around m^*, $\lambda = \upsilon = \theta_3$ there, and each expression is corrected by the integral of the gap over that interval.

[107] $\theta_3 \to +\infty$ as $p \to 0$ and $\theta_2 \to +\infty$ as $p \to \infty$, so υ exceeds the top of the support at both ends; and both are convex in p by Assumption 2, so $\{p : \upsilon(m) \leq 1\}$ is an interval for each m.

moves outwards, so economic departures become rarer, and the mobile dissenters who would otherwise leave are voicing instead. Both are exhausted once the loyalty ceiling has fallen far enough that political exit dominates. A flat or declining exit response early in an autocratisation episode is therefore a prediction of the model rather than evidence against it.[108]

Finally, that loyalty converges to the aligned population and that m^* alone divides the opponents hold for any distribution; the values $1/2$, $(1 + m^*)/4$ and $(1 - m^*)/4$ rely on the uniform benchmark, disagreement having median zero and being independent of mobility.[109]

[108] These shares are stocks, whereas departures and changes of agenda are observed as flows, so the empirical counterpart is closer to dS_k/dp than to S_k: for exit the prediction is a change of sign, not a difference in level.

[109] If the median researcher is an opponent, loyalty converges below one half and both silent responses are larger. If disagreement and mobility are positively associated, exit gains at every pressure what self-censorship loses and the dissenting half tilts towards departure, while the loyalty share, the endpoints at $p = 0$ and the pressures $\underline{p}$, p_E, $\bar{p}$ stay put. The shape of Figure 1.D.3 is robust to both; only the heights of the exit and self-censorship curves change.

1. Measuring Research Similarity

We measure changes in research direction using text similarity between publications. The title of each publication is encoded as a sentence embedding: a vector representation in which texts with similar meaning lie close to each other. We use Multilingual-E5-large (Wang et al., 2024), a transformer-based model from the BERT family (Devlin et al., 2019) trained on multilingual data, which allows us to embed English, Hungarian, Slovenian and Croatian texts in a common space.[110] The similarity between two publications is the cosine similarity between their embeddings. Our main measure is based on the titles of the publications, since abstracts are unevenly available across sources, languages and periods. Before computing similarities, we remove generic titles (e.g. "Preface", "Book review") and translated duplicates of the same publication.

Three features of embedding similarities deserve emphasis. First, raw similarity levels are not directly interpretable. Although cosine similarity can in principle range from -1 to 1, embeddings produced by models such as Multilingual-E5 occupy a narrow region of the vector space, so that even unrelated texts typically have a high similarity. In our data, the average similarity between two publications in the same field is about 0.81, with a standard deviation of about 0.03 (Table 1.E.1). Differences that look small in raw terms can therefore be meaningful. Baseline levels also differ across disciplines, for instance because titles are more formulaic in some disciplines than in others. We therefore compare publications within the same domain, field or subfield, so that differences in baseline similarity across disciplines do not drive our measures. Moreover, we never interpret similarity levels on their own: we compare changes in similarity for allies and opponents with changes for neutral researchers.

Second, the model maps texts in different languages into a common space, but this alignment is imperfect: two publications on the same topic are more similar when written in the same language than when written in two different languages. Table 1.E.1 illustrates this: average similarities are around 0.84 between Hungarian-language publications and around 0.80 between English-language publications. This matters for our analysis (especially for the cross-country comparisons) because researchers in Hungary increasingly publish in Hungarian after 2010. An author who published mainly in English before 2010 and mainly in Hungarian afterwards would mechanically appear less similar to their past work, even if the topics they study had not changed, which would mimic a change in research direction. To separate changes in content from changes in language, we also compute all similarity measures separately within each language. We thus compare English-language publications only with English-language publications and Hungarian-language publications only with Hungarian-language publications.

Third, the minimum and maximum similarities depend mechanically on the number of publications being compared. An author with n_t publications in year t and n_p publications in the reference period contributes $n_t \times n_p$ pairs. The more pairs there are, the more likely it is that at least one of them is very similar, which raises the maximum, and that at least one is very dissimilar, which lowers the minimum, even if the author's research direction is unchanged. Since researchers perceived as opponents tend to publish less after 2010, their maximum similarity to past work would fall for this reason alone. We therefore use the average similarity as our main measure, which does not depend on the total number of pairs.

[110] Publications in other languages are also embedded but are not treated as a separate category.

2. Similarity to one's own past research

For each author and year, we compute the pairwise similarities between the author's publications that year and their publications in a reference period, excluding papers that belong to both. Our baseline reference period is all publications before 2005, leaving the period 2005-2009 to be used for pre-trend checks before Orbán's election. As robustness checks, we use 2000–2005, 2005–2009 and all publications before 2009, and the most recent prior year in which the author published. We summarise each matrix by its mean, minimum and maximum. A decline in similarity to one's own pre-2010 work, relative to neutral researchers, thus indicates in our analysis a change in research direction.

3. Similarity to allies and opponents

For each author and year, we also compute the pairwise similarities between the author's publications and the publications of allies and of opponents in the same domain, field or subfield. If the author is a member of the group, we exclude the author's own publications. Our baseline reference period is again 2000–2009. For robustness, we use additional measures, taking 2000–2005, 2005–2009, all publications before 2005 and before 2009 as reference periods. As above, we summarise each matrix by its mean, minimum and maximum, using titles in our main specification and abstracts as a robustness check, both for all publications and separately for English and Hungarian-language publications. Thus, moving away from opponents' pre-2010 research is consistent with self-censorship. Moving toward allies' research is consistent with the *loyalty* strategy in the sense of Hirschman (1972) and as discussed in Section VI.

4. Descriptive statistics

Table 1.E.1 reports the distribution of pairwise similarities between publications of authors in the same field over 2000–2009, by pair of groups and language. Before 2010, allies, opponents and neutral researchers publish on very similar topics: across all languages, average similarities differ by at most 0.004 across pairs of groups, about a tenth of a standard deviation. The only noticeable difference concerns English-language publications, where opponents' publications are less similar to those of neutral researchers (0.799) than allies' publications are (0.812), a gap of about 0.4 standard deviations. No such gap appears for Hungarian-language publications.

Table 1.E.1 : Pairwise Similarity of Publications Across Groups, 2000–2009

	All languages				English				Hungarian			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Allies (within)	0.813	0.035	0.651	1	0.813	0.031	0.666	1	0.838	0.029	0.680	1
Opponents (within)	0.817	0.038	0.621	1	0.806	0.035	0.653	1	0.841	0.029	0.645	1
Allies – opponents	0.816	0.037	0.643	1	0.805	0.034	0.656	1	0.836	0.029	0.660	1
Allies – neutral	0.814	0.035	0.602	1	0.812	0.029	0.650	1	0.837	0.029	0.631	1
Opponents – neutral	0.816	0.037	0.608	1	0.799	0.034	0.637	1	0.837	0.029	0.635	1

Notes: The table reports the mean, standard deviation, minimum and maximum of the cosine similarity between Multilingual-E5-large embeddings of publications, for pairs of publications by authors in the same field published between 2000 and 2009. "Within" rows compare members of the same group with each other; other rows compare members of two different groups. The first panel includes publications in all languages, the second only publications in English, and the third only publications in Hungarian. The maximum equals 1 by construction: a publication co-authored by members of both sets being compared appears on both sides of the comparison, and its similarity with itself is exactly 1.

5. Publication Quality and Journals

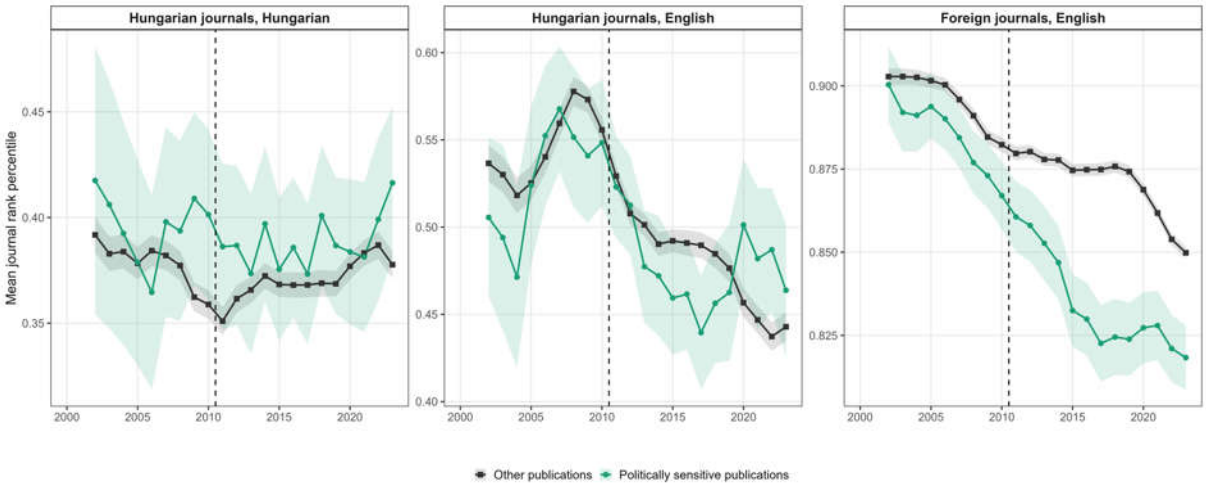
The decline in politically sensitive research documented above could come from two channels. On the *demand side*, journals, and in particular Hungarian journals whose editors, boards, and funding may be exposed to domestic political pressure, could have become less willing to publish sensitive work after 2010. On the *supply side*, researchers themselves could have stopped writing or submitting such work, which would reflect self-censorship. We provide descriptive evidence that the latter mechanism is at play, consistent with the results in Section VI. We classify as "politically sensitive" all publications that we identify as belonging to *gender* studies or to *sensitive* topics, as defined in Appendix 1.C.4, as well as publications on *minorities* and the *rule of law*.

We test two implications that distinguish the channels. First, *journal rank*: if editors became more hostile to sensitive work, sensitive papers should be pushed toward lower-ranked outlets after 2010, especially in Hungarian journals. If the shift comes from authors, the rank of sensitive and other publications should instead evolve similarly. Second, *journal-level persistence*: gatekeeping should change the editorial profile of specific journals, with some outlets ceasing to publish sensitive work. If the shift comes from authors, journals' propensity to publish sensitive work should remain stable, and the decline should be broad-based, including in foreign journals beyond the reach of Hungarian politics.

a) Journal Rank

Figure 1.E.1 plots the mean journal rank percentile of sensitive and other publications by Hungarian-affiliated researchers, separately for three types of outlets. In Hungarian-language Hungarian journals, where political pressure on editors would be most likely to operate, sensitive publications appear in outlets of similar or slightly higher rank than other publications throughout the period, with no break in the gap around 2010. In English-language Hungarian journals, both series rise until 2008 and then decline together: the post-2010 fall is common to all publications and not specific to sensitive work.

Figure 1.E.1 : Journal Rank of Politically Sensitive and Other Publications



Notes: The graph displays the mean journal rank percentile of publications by Hungarian-affiliated researchers, by year of publication, for politically sensitive publications (gender and sensitive topics, publications on minorities and the rule of law) and all other publications. Shaded areas represent 95% confidence intervals. The panels restrict to Hungarian journals publishing in Hungarian, Hungarian journals publishing in English, and foreign journals publishing in English. The dashed vertical line marks the 2010 change in government.

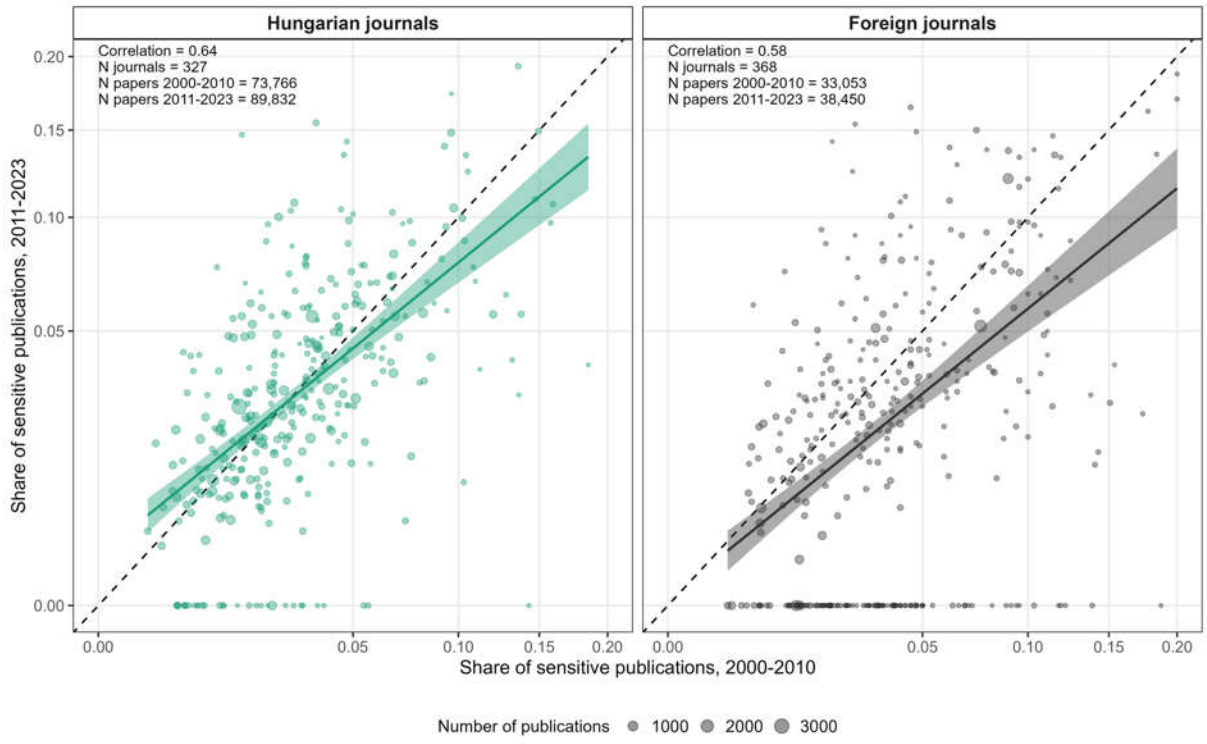
In foreign journals, the rank of sensitive publications declines more sharply than the that of other publications. This however is not likely to indicate an higher rejection rate of foreign journal specific to this type of publications due to the arrival of the illiberal government in 2010.

First, the gap opens around 2003, well before 2010, and does not accelerate afterwards. Second, editors of foreign journals do not have an obvious reason to respond to Hungarian domestic politics and political pressure. Third, the divergence is consistent with authors moving sensitive work from Hungarian-language outlets to English-language ones, where the audience is less exposed to domestic scrutiny, often in less prestigious international or regional journals. This compositional shift lowers the mean rank of sensitive publications in foreign journals without any change in editorial behaviour, and to the extent it operates after 2010, it reflects a language strategy by authors.

b) Journal-level Persistence

Figure 1.E.2 plots, for each journal, the share of sensitive publications among papers by Hungarian-affiliated researchers in 2011–2023 against the same share in 2000–2010. Journals' propensity to publish sensitive work appears to be highly persistent, with correlations of 0.64 for Hungarian and 0.58 for foreign journals. Their editorial orientation does not seem to change markedly. The decline is also not specific to Hungarian journals. Among Hungarian journals, the fitted line crosses the 45-degree line at a low pre-2010 share, consistent with reallocation across outlets rather than the closing of specific ones. Among foreign journals, which are not subject to Hungarian political pressure, the fitted line lies below the 45-degree line over almost the entire range, indicating a broad-based decline. Journals with no sensitive publications after 2011, mostly foreign, typically publish few papers by Hungarian researchers, so their zero share likely reflects small numbers of publications by Hungarian researchers rather than a change in the editorial policy.

Figure 1.E.2 : Share of Politically Sensitive Publications by Journal, Before and After 2010



Notes: Each dot represents a journal. The axes show the share of politically sensitive publications among publications by Hungarian-affiliated researchers in the journal over 2000–2010 (horizontal) and 2011–2023 (vertical). Dot size is proportional to the number of publications. The solid line is a linear fit with its 95% confidence interval. The dashed line represents the 45-degree line. The left panel covers 327 Hungarian journals (73,766 papers for the period 2000–2010, 89,832 in 2011–2023). The right panel covers 368 foreign journals (33,053 and 38,450 papers respectively).

Because journal-level shares are measured with noise, regression to the mean partly flattens the fitted lines relative to the 45-degree line, which however should not affect our conclusions. Regression to the mean leaves the average share unchanged, so the downward shift for foreign journals reflects a genuine decline. The same noise attenuates the correlations between the two periods, which are therefore lower bounds on the true persistence of journals' editorial profiles.

1.F — Political Alignment and Impact on Productivity: Complementary Results

1. Political Opponent, Political Ally

Table 1.F.1 : Effect on Academic Productivity - Allies

	Publications (1)	English (2)	Hungarian (3)	Top 10% Pub. (4)	Top 5% Pub. (5)
Allies × Post	0.135	0.117	0.182	0.048	-0.056
	(0.087)	(0.092)	(0.112)	(0.149)	(0.182)
Mean.Dep.Var.Pre	9	3.96	4.77	2.14	1.58
Convergence	TRUE	TRUE	TRUE	FALSE	TRUE
Observations	4,488	4,236	4,343	2,809	2,602
Author fixed effects	✓	✓	✓	✓	✓
Year-Entry year fixed effects	✓	✓	✓	✓	✓
Year-Field fixed effects	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those identified as loyal to the government. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total publications, publications in English, publications in Hungarian, and publications in the top 10% and the top 5% most cited journals in OpenAlex. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.2 : Effect on Academic Productivity - Political Opponents

	Publications (1)	English (2)	Hungarian (3)	Top 10% Pub. (4)	Top 5% Pub. (5)
Opponents × Post	-0.199***	-0.354***	-0.265***	-0.896**	-1.10**
	(0.069)	(0.101)	(0.081)	(0.374)	(0.520)
Mean.Dep.Var.Pre	3.47	1.41	1.80	0.193	0.162
Observations	10,981	8,836	10,192	1,359	721
Author fixed effects	✓	✓	✓	✓	✓
Year-Entry year fixed effects	✓	✓	✓	✓	✓
Year-Field fixed effects	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those identified as political opponents to the government. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total publications, publications in English, publications in Hungarian, and publications in the top 10% and the top 5% most cited journals in OpenAlex. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.3 : Effect on Research Quantity and Network – Political Opponents

	Panel A: Research Quantity			*Panel B: Research Network*	
	publications	pub. nat. language	pub. in english	nr co-authors	nr new co-authors
	(1)	(2)	(3)	(4)	(5)
pol. opp. × post (1rst mandate)	-0.3673*	-0.2525**	-0.0549	-0.3165**	-0.2590***
	(0.1887)	(0.1256)	(0.1009)	(0.1279)	(0.0863)
pol. opp. × post (2nd mandate)	-0.9105***	-0.6275***	-0.2015	-0.4885***	-0.3757***
	(0.2292)	(0.1557)	(0.1244)	(0.1536)	(0.1033)
pol. opp. × post (3rd mandate)	-1.171***	-0.6671***	-0.3934***	-0.6879***	-0.5647***
	(0.2883)	(0.2006)	(0.1360)	(0.1677)	(0.0982)
pol. opp. × post (4th mandate)	-2.015***	-0.9859***	-0.7947***	-0.8249***	-0.6440***
	(0.3106)	(0.2218)	(0.1535)	(0.2148)	(0.1311)
Observations	123,048	123,048	123,048	123,048	123,048
R^2	0.595	0.577	0.549	0.508	0.329
Within R^2	0.001	0.001	0.001	0.000	0.001
Mean.Dep.Var.Pre	3.094	2.206	0.539	1.345	0.888
author fixed effects	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions on data from national repositories. The unit of observation is author by year and the sample period is 2005–2023. Panel A reports the effect on research quantity (total publications, publications in national language, and publications in English); Panel B reports the effect on the author's research network (number of co-authors and number of new co-authors). Each column reports the decomposition of the treatment effect by political mandate. Control variables account for trends for the matching cell over time. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.4 : Effect on Research Quality – Political Opponents

	Top 1% Pub.		Top 5% Pub.		Bottom 50% Pub.	
	(1)	(2)	(3)	(4)	(5)	(6)
political opponent × post	-0.0100		-0.0699*		-0.0402	
	(0.0143)		(0.0384)		(0.0323)	
political opponent × post (1rst mandate)		-0.0053		-0.0717**		-0.0430
		(0.0144)		(0.0325)		(0.0328)
political opponent × post (2nd mandate)		-0.0086		-0.0710*		0.0017
		(0.0165)		(0.0424)		(0.0382)
political opponent × post (3rd mandate)		-0.0351**		-0.1188***		-0.0892**
		(0.0155)		(0.0460)		(0.0367)
political opponent × post (4th mandate)		-0.0598***		-0.1093**		-0.0712
		(0.0161)		(0.0482)		(0.0492)
Observations	123,048	123,048	123,048	123,048	123,048	123,048
R^2	0.343	0.343	0.567	0.567	0.422	0.422
Within R^2	0.000	0.001	0.000	0.001	0.000	0.000
Mean.Dep.Var.Pre	0.014	0.014	0.054	0.054	0.149	0.149
author fixed effects	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓

Note: results are from OLS Fixed-Effect regression, for each outcome, on data from Hungarian national repositories. The unit of observation is author by year and the sample period is from 2005-2024. The dependent variable is the number of publications in the top 1% and 5% most cited journals and the 50% least cited journals (resp. columns (1), (2), (3)), and the share of the latter over the total number of publications (column (4)). Standard errors (SE) are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

2. Impact on the Members of the Networks of Politically Aligned Researchers

Table 1.F.5 : Effect on Academic Productivity - Network of Political Opponents

Outcome	Research Quantity						Research Quality			
	publications		pub. in english		pub. nat. language		Top 10% Pub.		Top 5% Pub.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Network pol. opp. × post	-0.032		-0.015		-0.017		-0.125*		-0.092	
	(0.029)		(0.042)		(0.032)		(0.068)		(0.079)	
Network pol. opp. × post (1rst m)		-0.007		9.1×10^{-5}		0.024		-0.046		0.015
		(0.026)		(0.036)		(0.030)		(0.067)		(0.082)
Network pol. opp. × post (2nd m)		-0.082**		-0.076*		-0.033		-0.253***		-0.238***
		(0.033)		(0.045)		(0.039)		(0.066)		(0.077)
Network pol. opp. × post (3rd m)		-0.066*		-0.053		-0.048		-0.164*		-0.169*
		(0.038)		(0.051)		(0.048)		(0.084)		(0.096)
Network pol. opp. × post (4th m)		-0.042		0.010		-0.088		-0.138		-0.156
		(0.054)		(0.071)		(0.068)		(0.107)		(0.136)
Mean dep. var. (pre)	4.857	4.857	2.344	2.344	2.915	2.915	0.822	0.822	0.579	0.579
Observations	75,449	75,449	70,931	70,931	70,527	70,527	45,464	45,464	40,198	40,198
author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as political opponents is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. Panel A reports the effect on research quantity: total publications (columns 1-2), publications in English (columns 3-4), publications in Hungarian (columns 5-6). Panel B reports the effect on -research quality: publications in the top 10% (columns 7-8) and the top 5% most cited journals in OpenAlex (columns 9-10). The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.6 : Effect on Academic Productivity - Network of Allies

Outcome	Research Quantity						Research Quality			
	publications		pub. in english		pub. nat. language		Top 10% Pub.		Top 5% Pub.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Network ally × post	0.081*		0.077		0.021		0.144*		0.166*	
	(0.045)		(0.059)		(0.051)		(0.086)		(0.100)	
Network ally × post (1rst m)		0.067		0.070		-0.006		0.198**		0.292***
		(0.041)		(0.054)		(0.048)		(0.081)		(0.099)
Network ally × post (2nd m)		0.021		-0.005		-0.034		0.095		0.144
		(0.059)		(0.075)		(0.065)		(0.092)		(0.110)
Network ally × post (3rd m)		0.065		0.011		0.112		0.115		0.084
		(0.056)		(0.070)		(0.073)		(0.110)		(0.139)
Network ally × post (4th m)		0.131*		0.147*		0.128		0.176		0.088
		(0.074)		(0.083)		(0.113)		(0.137)		(0.169)
Mean dep. var. (pre)	4.660	4.660	2.536	2.536	2.614	2.614	0.884	0.884	0.639	0.639
Observations	31,085	31,085	29,648	29,648	28,553	28,553	22,280	22,280	19,850	19,850
author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as loyal to the government is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. Panel A reports the effect on research quantity: total publications (columns 1-2), publications in English (columns 3-4), publications in Hungarian (columns 5-6). Panel B reports the effect on -research quality: publications in the top 10% (columns 7-8) and the top 5% most cited journals in OpenAlex (columns 9-10). The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

3. Targeted Topics: Climate

The effect on researchers specialised in climate topics is not the same as the one we estimate on researchers specialised in other politically sensitive topics: Table 1.F.7 shows that although the effect on publications is negative, the effect on quality appears to be positive. Our hypothesis is that defiance towards climate science is less pronounced in Orbán's regime than in other contemporary right-wing illiberal regimes.

Table 1.F.7 : Effect on Academic Productivity - Climate

	Publications		count_pub_top_1pct	count_pub_top_10pct	citation_count	count_new_collaboration
	(1)	(2)	(3)	(4)	(5)	(6)
post × climate	-0.098***		0.331**			
	(0.037)		(0.159)			
post (1rst m) × climate		-0.068**		0.030	0.086	0.032
		(0.031)		(0.071)	(0.067)	(0.051)
post (2nd m) × climate		-0.092**		0.070	0.168**	-0.039
		(0.044)		(0.084)	(0.074)	(0.062)
post (3rd m) × climate		-0.087		0.105	0.292***	-0.042
		(0.057)		(0.096)	(0.097)	(0.070)
post (4th m) × climate		-0.100		0.240**	0.220	-0.022
		(0.072)		(0.116)	(0.144)	(0.101)
Mean dep. var. (pre)	6.081	6.081	0.504	0.504	38.290	3.112
Observations	40,324	40,324	10,211	215,173	38,687	39,762
Author fixed effects	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those who published at least two publications in gender topics for the first line and sensitive topics for the second line. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total publications (columns 1-2), publications in English (columns 3-4), publications in Hungarian (columns 5-6) and publications in the top 5% most cited journals in OpenAlex. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

4. Targeted Topics for Political Opponents

In this paper, we focus on separating the effect of being perceived as a political opponent and working on topics targeted by the government, excluding researchers who cumulate both. In this subsection, we analyse the effect of illiberal policies on this subset of researchers. These results cannot be interpreted causally, as there are only very few of these researchers who belong to the group of perceived opponents and the group of researchers working in gender (17 researchers) or sensitive topics (79 researchers).

In order to assess whether they suffer an additional penalty, we estimate a triple difference-in-differences regression in which we compare neutral researchers to perceived political opponents, researchers working in targeted topics, and perceived political opponents working on gender topics. The regression is estimated on matched data in which we match neutral researchers to researchers receiving any value of treatment following the same matching procedure as in Section IV.2.

Table 1.F.8 and Table 1.F.9 show the results of this estimation. While coefficients are no longer significant for the political opponents treatment due to the small number of units belonging to the group in which treatments are interacted, the effect of the interaction dwarfs any other effect for both groups of targeted topics. These tables confirm that cumulating both reasons for political pressures increases the negative effect suffered by targeted researchers.

5. Channels of repression

a) Funding and the Effect of Illiberal Policies

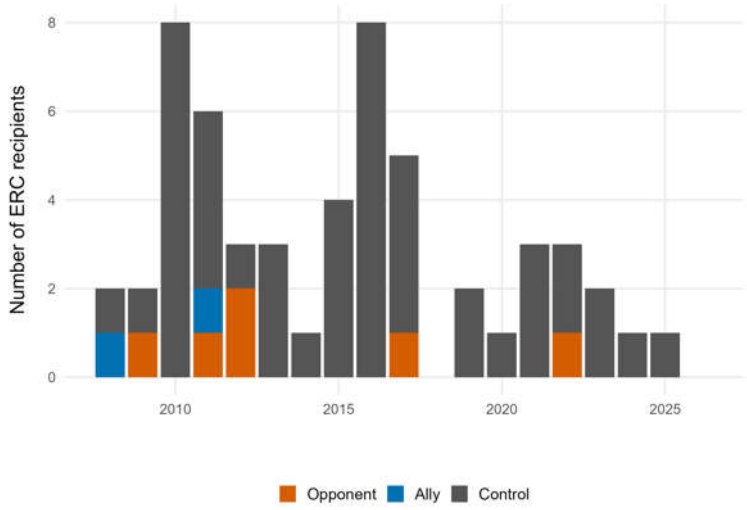
ERC Funding

Figure 1.F.1 shows the total number of ERC recipients per year among our sample of political opponents, allies and researchers from the database tagged as neutral. There does not appear to be a strong decreasing trend in total number of ERC recipients post-2010. The composition of the groups is quite different: most political opponents who received an ERC grant work in social sciences. Restricting the focus to researchers in social sciences, Figure 1.F.2 shows a decreasing trend in the average number of publications per year for ERC recipients among political opponents. Table 1.F.10 and Table 1.F.11 report the regression results discussed in Section V.2.a.

Differential Effects Between Public and Private Affiliations

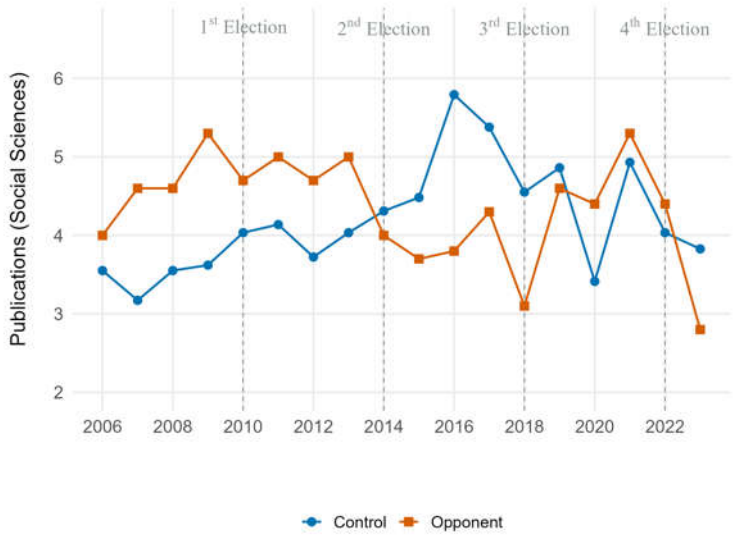
The main results are reported in Table 1.2 in the main text. Table 1.F.12 reports a placebo exercise on researchers who are neither political opponents, allies, nor co-authors of either group: public affiliation before 2010 is not associated with a post-2010 break for any outcome.

Figure 1.F.1 : Number of ERC recipients per year per political status



Notes: The graph above reports the total number of ERC recipients by call year for political opponents and for other Hungarian researchers for the period 2006-2023.

Figure 1.F.2 : Number of Publications for ERC Recipients, Social Sciences, per Political Group



Notes: The graph above reports the average number of publications for ERC recipients in social sciences each year, for political opponents and for other Hungarian researchers for the period 2006-2023.

Table 1.F.8 : Trends in Outcomes for Political Opponents, Gender-Related Topics and Political Opponents in Gender Topics

	Publications		Publications (EN)		Publications (HU)		Top 5% Pub.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
pol. opp. × post	-0.084		-0.092		-0.111*		-0.248	
	(0.051)		(0.084)		(0.063)		(0.226)	
gender × post	-0.059*		-0.079**		-0.057		-0.171**	
	(0.031)		(0.038)		(0.041)		(0.072)	
pol. opp. × gender × post	0.012		-0.253		0.094		1.05**	
	(0.153)		(0.176)		(0.200)		(0.523)	
pol. opp. × post (1rst m)		-0.029		-0.090		-0.020		-0.345**
		(0.045)		(0.077)		(0.056)		(0.167)
pol. opp. × post (2nd m)		-0.110*		-0.104		-0.144*		-0.072
		(0.060)		(0.094)		(0.080)		(0.239)
pol. opp. × post (3rd m)		-0.103		-0.047		-0.161*		-0.264
		(0.070)		(0.112)		(0.091)		(0.415)
pol. opp. × post (4th m)		-0.187*		-0.259*		-0.246**		-0.740
		(0.098)		(0.145)		(0.111)		(0.553)
gender × post (1rst m)		-0.033		-0.027		-0.026		-0.091
		(0.027)		(0.034)		(0.037)		(0.071)
gender × post (2nd m)		-0.031		-0.058		-0.019		-0.197**
		(0.037)		(0.045)		(0.050)		(0.086)
gender × post (3rd m)		-0.123***		-0.162***		-0.138**		-0.271**
		(0.044)		(0.053)		(0.059)		(0.105)
gender × post (4th m)		-0.084		-0.137**		-0.092		-0.250*
		(0.054)		(0.067)		(0.078)		(0.141)
pol. opp. × gender × post (1rst m)		0.079		-0.029		0.179		1.42***
		(0.124)		(0.182)		(0.198)		(0.419)
pol. opp. × gender × post (2nd m)		0.092		-0.261		0.035		1.01
		(0.182)		(0.225)		(0.232)		(0.725)
pol. opp. × gender × post (3rd m)		-0.080		-0.447		0.154		0.610
		(0.266)		(0.278)		(0.314)		(0.816)
pol. opp. × gender × post (4th m)		-0.792		-1.81***		-0.584		-8.27***
		(0.577)		(0.690)		(0.595)		(0.631)
Mean dep. var. (pre)	4.239	4.239	1.501	1.501	3.114	3.114	0.307	0.307
Observations	129,986	129,986	111,002	111,002	123,815	123,815	38,972	38,972
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect PPML triple difference-in-differences regressions, for each outcome, on data from MTMT. The unit of observation is author by year and the sample period is from 2005-2024. Treated authors are authors from our sample of political opponents, split by whether they also belong to the group of researchers working on gender-related topics, and researchers working on gender-related topics. Control authors are neutral Hungarian researchers matched to researchers receiving either or both treatment. We report the effect on total publications, publications in English, publications in Hungarian, citations, and publications in the top 5% most cited journals in OpenAlex. Control variables account for trends for field and entry year in academia FE. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.9 : Trends in Outcomes for Political Opponents, Sensitive Topics and Political Opponents in Sensitive Topics

	Publications		Publications (EN)		Publications (HU)		Top 5% Pub.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
pol. opp. × post	0.057		0.086		0.014		-0.070	
	(0.056)		(0.091)		(0.072)		(0.363)	
sensitive × post	-0.055*		0.018		-0.102**		-0.358**	
	(0.028)		(0.044)		(0.047)		(0.169)	
pol. opp. × sensitive × post	-0.287***		-0.526***		-0.166		-0.272	
	(0.094)		(0.144)		(0.123)		(0.524)	
pol. opp. × post (1rst m)		0.049		0.075		0.020		-0.082
		(0.051)		(0.074)		(0.067)		(0.308)
pol. opp. × post (2nd m)		0.032		0.091		-0.018		0.077
		(0.063)		(0.101)		(0.087)		(0.385)
pol. opp. × post (3rd m)		0.094		0.130		0.050		-0.188
		(0.075)		(0.126)		(0.106)		(0.496)
pol. opp. × post (4th m)		0.050		-0.099		-0.022		-0.373
		(0.109)		(0.171)		(0.127)		(0.670)
sensitive × post (1rst m)		-0.041*		0.005		-0.070**		-0.441***
		(0.024)		(0.040)		(0.035)		(0.166)
sensitive × post (2nd m)		-0.068**		0.028		-0.132**		-0.392**
		(0.034)		(0.050)		(0.056)		(0.191)
sensitive × post (3rd m)		-0.068*		0.019		-0.109		-0.275
		(0.041)		(0.061)		(0.072)		(0.217)
sensitive × post (4th m)		-0.003		0.031		-0.092		0.052
		(0.053)		(0.078)		(0.082)		(0.297)
pol. opp. × sensitive × post (1rst m)		-0.127		-0.366***		0.022		-0.467
		(0.086)		(0.140)		(0.113)		(0.584)
pol. opp. × sensitive × post (2nd m)		-0.264**		-0.574***		-0.149		0.094
		(0.114)		(0.169)		(0.154)		(0.553)
pol. opp. × sensitive × post (3rd m)		-0.458***		-0.647***		-0.385**		-0.456
		(0.134)		(0.205)		(0.180)		(0.759)
pol. opp. × sensitive × post (4th m)		-0.769***		-0.716**		-0.576**		-1.26
		(0.202)		(0.301)		(0.229)		(1.18)
Mean dep. var. (pre)	4.052	4.052	1.160	1.160	3.188	3.188	0.202	0.202
Observations	134,047	134,047	110,214	110,214	128,640	128,640	23,768	23,768
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect PPML triple difference-in-differences regressions, for each outcome, on data from MTMT. The unit of observation is author by year and the sample period is from 2005-2024. Treated authors are authors from our sample of political opponents, split by whether they also belong to the group of researchers working on sensitive topics, and researchers working on sensitive topics. Control authors are neutral Hungarian researchers matched to researchers receiving either or both treatment. We report the effect on total publications, publications in English, publications in Hungarian, citations, and publications in the top 5% most cited journals in OpenAlex. Control variables account for trends for field and entry year in academia FE. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.10 : Trends in Outcomes for Political Opponents, by closeness to pre-2010 ERC recipients

	publications	pub. in english	pub. nat. language	citation_count	nr pub. top 5%
	(1)	(2)	(3)	(4)	(5)
pol. opp × ERC exposure × post	0.4556	0.2633**	0.0775	-24.42***	-0.0478**
	(0.3482)	(0.1317)	(0.3294)	(4.111)	(0.0213)
ERC exposure × post	0.5017***	0.2029***	0.4662***	-5.856***	0.0190*
	(0.1873)	(0.0709)	(0.1772)	(2.212)	(0.0115)
pol. opp × post	-0.5565***	-0.1363***	-0.4322***	-2.091	0.0052
	(0.1380)	(0.0522)	(0.1305)	(1.629)	(0.0084)
Observations	21,600	21,600	21,600	21,600	21,600
R^2	0.599	0.487	0.599	0.355	0.351
Within R^2	0.002	0.001	0.001	0.005	0.000
Mean.Dep.Var.Pre	3.538	0.575	2.992	13.878	0.025
author fixed effects	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect PPML regressions of equation (1.2), for each outcome, on data from Hungarian national repository. The unit of observation is author by year and the sample period is from 2005-2024. Treated authors are authors from our sample of political opponents, split by whether their collaboration index with pre-2010 ERC recipients is above or below median. Control authors are neutral Hungarian researchers matched to the treated, split whether their collaboration index with pre-2010 ERC recipients is above or below median. We report the effect on total publications, publications in English, publications in Hungarian, citations count and publications in the top 10% most cited journals. Control variables account for trends for field and entry year in academia FE. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.11 : Trends in Outcomes for Political Opponents, by closeness to any-period ERC recipients

	publications	pub. in english	pub. nat. language	citation_count	count_pub_top_10pct
	(1)	(2)	(3)	(4)	(5)
pol. opp × ERC exposure × post	0.1337**	0.4147***	0.0662	0.0050	0.0564
	(0.0618)	(0.1314)	(0.0718)	(0.0290)	(0.6039)
ERC exposure × post	0.1130***	0.0474	0.1190***	0.0109	-0.0625
	(0.0347)	(0.0770)	(0.0388)	(0.0153)	(0.4310)
pol. opp × post	-0.1668***	-0.2418***	-0.1980***	-0.1743***	-0.0890
	(0.0262)	(0.0592)	(0.0301)	(0.0156)	(0.3899)
Observations	21,339	17,064	20,259	17,209	2,460
Squared Correlation	0.607	0.504	0.602	0.623	0.658
Pseudo R^2	0.410	0.334	0.424	0.639	0.345
Mean.Dep.Var.Pre	3.275	1.073	2.367	24.189	0.330
author fixed effects	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓

Note: results are from Fixed Effect PPML regressions of equation (2), for each outcome, on data from Hungarian national repository. The unit of observation is author by year and the sample period is from 2005-2024. Treated authors are authors from our sample of political opponents, split by whether their collaboration index with researchers who received an ERC grant at any point is above or below median. Control authors are neutral Hungarian researchers matched to the treated, split whether their collaboration index with researchers who received an ERC grant at any point is above or below median. We report the effect on total publications, publications in English, publications in Hungarian, citations and publications in the top 10% most cited journals. Control variables account for trends for field and entry year in academia FE. Since grants awarded after 2010 depend on post-2010 productivity, these results are descriptive. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.12 : Placebo: Public Affiliation among Researchers Unrelated to Opponents or Allies

	publications (1)	pub. in english (2)	pub. nat. language (3)	citation (nat) (4)	nr pub. top 5% (5)
public × post	-0.0412	-0.0863	0.0167	0.0113	-0.1071
	(0.0403)	(0.0574)	(0.0446)	(0.1177)	(0.1308)
Observations	80,874	78,505	78,167	79,869	43,600
Squared Correlation	0.655	0.738	0.588	0.854	0.860
Pseudo R^2	0.427	0.487	0.441	0.752	0.499
Mean.Dep.Var.Pre	5.195	2.597	3.112	48.091	0.472
author fixed effects	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect PPML regressions, for each outcome, on data from the Hungarian national repository. The unit of observation is author by year and the sample period is from 2005-2024. The sample includes researchers who are neither political opponents, allies, nor co-authors of either group. The coefficient compares researchers affiliated to a public institution before 2010 with other researchers, after 2010. Control variables account for trends for field and entry year in academia FE. *** p<0.01, ** p<0.05, * p<0.1.

b) **International Network and Loss of Influence**

Table 1.F.13 : International Collaborations - Opponents

	Intl. collab.		Regional collab.		US collab.	
	(1)	(2)	(3)	(4)	(5)	(6)
pol. opp × post	0.128		1.81***		0.178*	
	(0.080)		(0.459)		(0.105)	
pol. opp. × post (1rst m)		-0.066		1.28***		0.202**
		(0.068)		(0.425)		(0.093)
pol. opp. × post (2nd m)		0.233***		2.55***		0.224**
		(0.085)		(0.589)		(0.107)
pol. opp. × post (3rd m)		0.299**		2.15***		0.297*
		(0.136)		(0.754)		(0.163)
pol. opp. × post (4th m)		0.222*		0.688		0.003
		(0.127)		(0.654)		(0.111)
Mean dep. var. (pre)	0.148	0.148	1.798	1.798	0.114	0.114
Observations	25,524	25,524	25,524	25,524	25,524	25,524
Author fixed effects	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions. The unit of observation is author by year. Treatment is being perceived as a political opponent. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on the number of publications with at least a co-author outside of Hungary in columns 1-2, in Slovenia, Slovakia or Czechia in columns 3-4, and in the USA in columns 5-6. Odd columns report the effect for the entire period; even columns report the effect by mandate. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.14 : International Collaborations - Allies

	Intl. collab.		Regional collab.		US collab.	
	(1)	(2)	(3)	(4)	(5)	(6)
ally × post	-0.418***		-1.12		-0.189	
	(0.157)		(0.757)		(0.159)	
ally × post (1rst m)		-0.382**		-0.582*		-0.215
		(0.154)		(0.325)		(0.143)
ally × post (2nd m)		-0.305		-1.28		-0.118
		(0.193)		(0.817)		(0.192)
ally × post (3rd m)		-0.594**		-1.96*		-0.264
		(0.248)		(1.11)		(0.229)
ally × post (4th m)		-0.552*		-2.87**		-0.051
		(0.289)		(1.16)		(0.284)
Mean dep. var. (pre)	0.686	0.686	3.182	3.182	0.444	0.444
Observations	31,734	31,734	31,734	31,734	31,734	31,734
Author fixed effects	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓

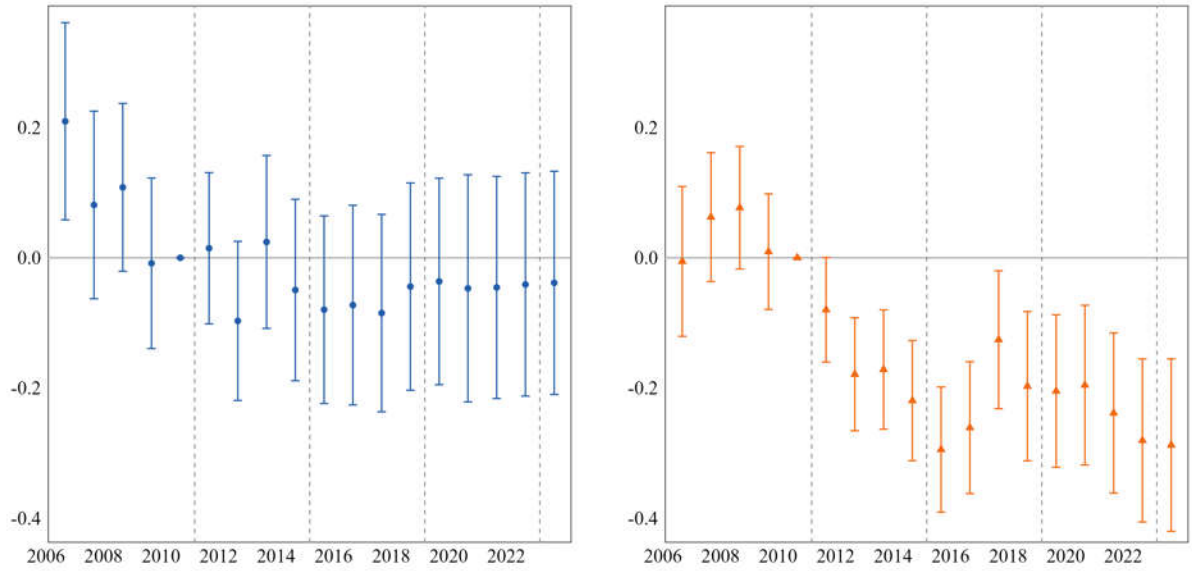
Note: Results are from Fixed-Effect OLS regressions. The unit of observation is author by year. Treated researchers are those that we identified as loyal to the government. We report the effect on the number of publications with at least a co-author outside of Hungary in columns 1-2, in Slovenia, Slovakia or Czechia in columns 3-4, and in the USA in columns 5-6. Odd columns report the effect for the entire period; even columns report the effect by mandate. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.F.15 : Network of Allies: Effect on Network and Influence

	Nr Coauthors		Nr New Collaborations		Citations	
	(1)	(2)	(3)	(4)	(5)	(6)
Network ally × post	-0.091*		-0.099*		-0.036	
	(0.048)		(0.055)		(0.064)	
Network ally × post (1rst m)		-0.090**		-0.142***		-0.016
		(0.041)		(0.054)		(0.060)
Network ally × post (2nd m)		-0.100*		-0.088		-0.025
		(0.053)		(0.062)		(0.075)
Network ally × post (3rd m)		-0.103*		-0.066		-0.082
		(0.062)		(0.069)		(0.090)
Network ally × post (4th m)		-0.060		-0.027		0.134
		(0.072)		(0.096)		(0.141)
Mean dep. var. (pre)	6.205	6.205	3.295	3.295	42.524	42.524
Observations	30,929	30,929	30,737	30,737	30,358	30,358
author fixed effects	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect PPML regressions on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as loyal to the government is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total number of co-authors (columns 1-2), total number of new co-authors (columns 3-4), and citations from other MTMT publications (columns 5-6). The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Odd columns report the average effect on the period 2010-2024 and even columns split the effect by mandate. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

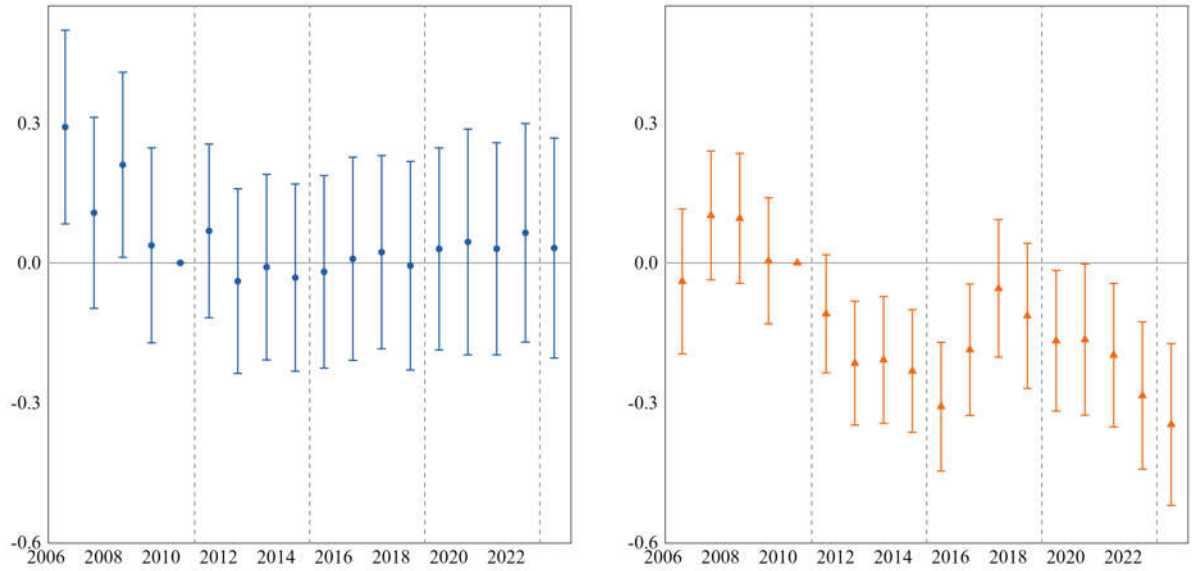
Figure 1.F.3 : Ally Network and Opponent Network: Effect on Total Collaborations



(a) Ally Network (b) Opponent Network

Note: Results are from Fixed-Effect PPML regressions from CRIS data. The unit of observation is author by year and the sample period is 2005–2023. The outcome variable is the total number of collaborations. Panel (a) reports the effect for the Ally Network individuals i.e., researchers which collaboration index with ally researchers is above the median. Panel (b) reports the equivalent estimation for opponent network. Control variables account for trends field and cohort dynamics over time. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

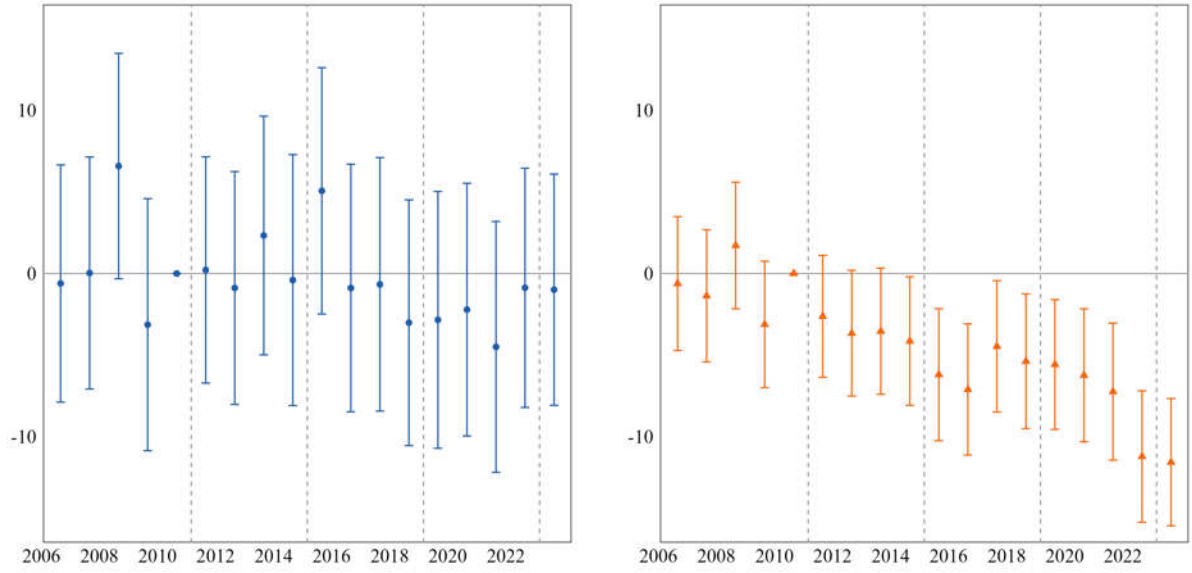
Figure 1.F.4 : Ally Network and Opponent Network: Effect on New Collaborations



(a) Ally Network (b) Opponent Network

Note: Results are from Fixed-Effect PPML regressions from CRIS data. The unit of observation is author by year and the sample period is 2005–2023. Panel (a) reports the effect for the Ally Network individuals i.e., researchers which collaboration index with ally researchers is above the median. Panel (b) reports the equivalent estimation for opponent network. Control variables account for trends field and cohort dynamics over time. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1.F.5 : Ally Network and Opponent Network: Effect on Citations



(a) Ally Network

(b) Opponent Network

Note: Results are from Fixed-Effect OLS regressions from CRIS data. The unit of observation is author by year and the sample period is 2005–2023. The outcome variable is the count of citations per year. Panel (a) reports the effect for the Ally Network individuals i.e., researchers which collaboration index with ally researchers is above the median. Panel (b) reports the equivalent estimation for opponent network. Control variables account for trends field and cohort dynamics over time. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

1. Voicing

Table 1.G.1 : Effect on *Voicing* behaviours – Political Opponents Treatment

	pub. *rule of law*			share pub. *rule of law*		
	(1)	(2)	(3)	(4)	(5)	(6)
pol. opp. × post	0.0775***			0.0153***		
	(0.0247)			(0.0041)		
CEU × post		0.0586**			0.0103**	
		(0.0253)			(0.0046)	
NGO × post			0.1632**			0.0271***
			(0.0724)			(0.0094)
Observations	123,048	121,932	120,852	78,546	77,675	76,936
R^2	0.311	0.289	0.300	0.243	0.238	0.229
Within R^2	0.002	0.001	0.002	0.001	0.000	0.001
Mean.Dep.Var.Pre	0.006	0.006	0.006	0.002	0.002	0.002
author fixed effects	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect OLS regression, for each outcome, on data from MTMT, the Hungarian CRIS-national repositories. The unit of observation is author by year and the sample period is from 2005-2024. Treated individuals are all researchers perceived as political opponents (line 1), researchers who were professors at the CEU prior to 2010 (line 2) and researchers who were named in the list published by Figyelő in 2018. Columns 1 to 3 report the effect on each group for the number of publications identified as mentioning the subject of rule of law; columns 4-6 report the effect on the share of publications mentioning the subject. Control variables account for trends for the matching cell over time. Standard errors (SE) are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

2. Ally Network

Table 1.G.2 : Ally Network: Migration and International Collaboration

	Migration				Collaboration					
	Any affil. abroad		Abroad (2y)		Intl. collab.		Regional collab.		Retiring	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Network ally × post	-0.004		-6.28×10^{-5}		-0.160		-0.635***		-0.0002	
	(0.011)		(0.008)		(0.181)		(0.190)		(0.0006)	
Network ally × post (1rst m)		-0.018		-0.013*		-0.045		-0.078**		-0.0008
		(0.011)		(0.007)		(0.226)		(0.033)		(0.0008)
Network ally × post (2nd m)		0.010		0.010		-0.184		-0.162***		-0.0006
		(0.013)		(0.010)		(0.252)		(0.043)		(0.0008)
Network ally × post (3rd m)		-0.006		0.002		-0.233		-0.155***		0.0003
		(0.014)		(0.010)		(0.158)		(0.055)		(0.0008)
Network ally × post (4th m)		0.005		0.003		-0.236**		-0.198***		0.001
		(0.020)		(0.011)		(0.095)		(0.063)		(0.001)
Mean dep. var. (pre)	0.195	0.195	0.091	0.091	0.663	0.663	3.475	3.475	0.360	0.360
Observations	146,632	146,632	239,670	239,670	239,670	239,670	239,670	230,480	239,670	239,670
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions on data from OpenAlex. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as loyal to the government is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. Panel A reports the effect on migration: probability of having an affiliation outside of Hungary for the past 2 years (columns 1-2), the past 3 years (columns 3-4), and the past 4 years (columns 5-6). Panel B reports the effect on number of publications with US researchers. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.G.3 : Ally Network: Additional Measures

	Migration						Collaboration	
	Abroad (recent)		Abroad (3y)		Abroad (4y)		US collab.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network ally × post	-0.006		0.003		0.006		-0.124	
	(0.008)		(0.008)		(0.007)		(0.168)	
Network ally × post (1^{rst} m)		-0.008		-0.007		-0.001		-0.043
		(0.008)		(0.007)		(0.006)		(0.211)
Network ally × post (2nd m)		-0.001		0.007		0.005		-0.167
		(0.009)		(0.009)		(0.008)		(0.235)
Network ally × post (3rd m)		-0.008		0.007		0.013		-0.166
		(0.010)		(0.010)		(0.009)		(0.140)
Network ally × post (4th m)		-0.005		0.005		0.013		-0.109
		(0.011)		(0.010)		(0.009)		(0.091)
Mean dep. var. (pre)	0.099	0.099	0.074	0.074	0.060	0.060	0.482	0.482
Observations	226,390	226,390	239,670	239,670	239,670	239,670	239,670	239,670
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions on data from OpenAlex. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as loyal to the government is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. Panel A reports the effect on migration excluding all non-publishing years: probability of having an affiliation outside of Hungary for the past 2 years (columns 1-2), the past 3 years (columns 3-4), and the past 4 years (columns 5-6). Panel B reports the effect on number of publications with US researchers. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

3. Political Opponent Network

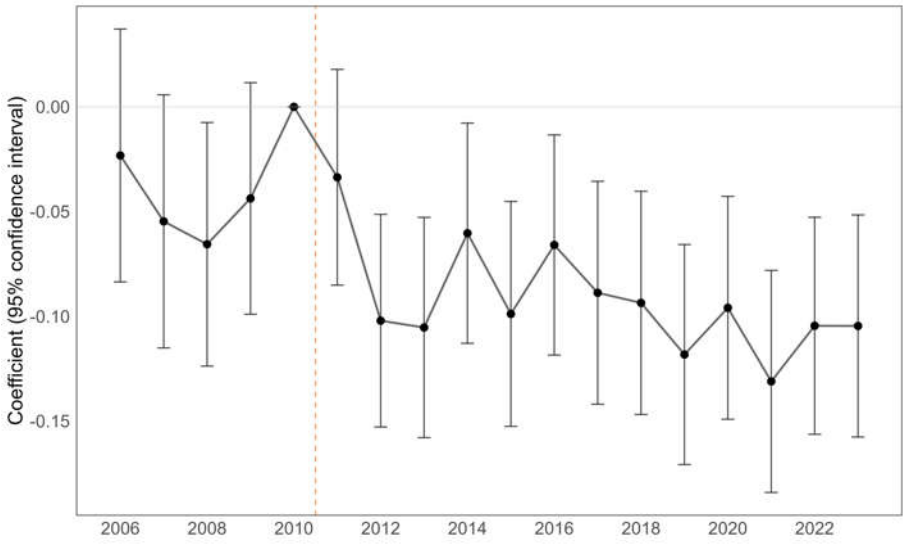
Table 1.G.4 : Political Opponent Network: Additional Measures

	Migration						Collaboration	
	Abroad (recent)		Abroad (3y)		Abroad (4y)		US collab.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network opp. × post	0.030*		0.024**		0.023**		-0.034	
	(0.018)		(0.011)		(0.011)		(0.181)	
Network pol. opp. × post (1rst m)		0.004		-0.003		-4.24×10^{-5}		0.037
		(0.015)		(0.009)		(0.009)		(0.199)
Network pol. opp. × post (2nd m)		0.030		0.024*		0.021*		0.098
		(0.020)		(0.012)		(0.012)		(0.198)
Network pol. opp. × post (3rd m)		0.053**		0.044***		0.042***		-0.228
		(0.023)		(0.016)		(0.016)		(0.237)
Network pol. opp. × post (4th m)		0.048*		0.049***		0.042**		-0.347
		(0.024)		(0.018)		(0.017)		(0.350)
Mean dep. var. (pre)	0.117	0.117	0.050	0.050	0.039	0.039	0.221	0.221
Observations	112,450	112,450	128,772	128,772	128,772	128,772	47,565	47,565
Author fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions on data from OpenAlex. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those whose collaboration index with researchers perceived as political opponents is above median. Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. Panel A reports the effect on migration excluding all non-publishing years: probability of having an affiliation outside of Hungary for the past 2 years (columns 1-2), the past 3 years (columns 3-4), and the past 4 years (columns 5-6). Panel B reports the effect on number of publications with US researchers. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

4. Additional Results on Research Similarity and Self-censorship

Figure 1.G.1 : Effect on the Probability of Publishing in Gender - Triple DiD



Note: Results are from Fixed-Effect OLS triple difference-in-differences regressions on data from the CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. The line reports the difference in probability of publishing any paper with gender topics, between researchers who publish in gender topics in Hungary compared to those publishing in gender topics in Slovenia or Croatia, compared to their respective difference in probability with researchers in their country who do not publish in gender topics. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.G.5 : Triple Difference-in-Differences: Researchers working in Gender

	Pubs. (1)	Co-authors (2)	Pub. Gender (3)	Publish in Gender (4)
Post × gender	-0.006	0.014	-0.215***	-0.151***
	(0.028)	(0.029)	(0.027)	(0.010)
Post × hungary × gender	-0.094***	-0.142***	-0.105**	-0.039***
	(0.036)	(0.040)	(0.043)	(0.014)
Observations	353,680	335,814	377,767	377,767
Author fixed effects	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓
Country × Cohort × Cohort fixed effects	✓	✓	✓	✓

Note: Results are from Fixed-Effect Triple difference-in-differences regressions on data from the CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those who published at least 3 publications in gender topics (first line) and those who do so in Hungary (second line). Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total publications, total number of co-authors, total publications with gender keywords in the title, and probability to publish any paper with gender keywords in its titles. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.G.6 : Triple Difference-in-Differences: Researchers working on Sensitive Topics

	Publications (1)	Co-authors (2)	Pub. Sensitive (3)	Publish in Sensitive (4)
sensitive × Post	0.015	0.038	-0.203***	-0.166***
	(0.035)	(0.050)	(0.028)	(0.014)
hungary × sensitive × Post	-0.113***	-0.311***	-0.035	0.001
	(0.042)	(0.066)	(0.037)	(0.017)
Observations	281,864	261,129	299,629	299,629
Author fixed effects	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓
Country × Cohort × Cohort fixed effects	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS triple difference-in-differences regressions on data from the CRIS-national repositories. The unit of observation is author by year and the sample period is 2005–2024. Treated authors are those who published at least 3 publications in sensitive topics (first line) and those who do so in Hungary (second line). Authors are matched using their number of publications over the period 2000-2005, entry year in academia and field of research. We report the effect on total publications, total number of co-authors, total publications with sensitive topic keywords in the title, and probability to publish any paper with sensitive topic keywords in its titles. The dataset is winsorised at the top of the distribution at the 1 percent level for the outcome variable. Control variables account for trends for field and entry year in academia FE. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.G.7 : Researchers Working on Sensitive Topics: Research Similarity

	sim. pre-2005 (HU)		sim. pre-2005 (EN)		sim. Ally	
	(1)	(2)	(3)	(4)	(5)	(6)
sensitive × post	-0.007***		0.001		-0.010***	
	(0.002)		(0.0007)		(0.002)	
sensitive × post (1rst m)		-0.006***		-8.19×10^{-5}		-0.003
		(0.002)		(0.0008)		(0.002)
sensitive × post (2nd m)		-0.007***		0.0007		-0.011***
		(0.002)		(0.0009)		(0.003)
sensitive × post (3rd m)		-0.011***		0.002**		-0.016***
		(0.003)		(0.0009)		(0.004)
sensitive × post (4th m)		-0.011***		0.002*		-0.014**
		(0.004)		(0.001)		(0.006)
Mean dep. var. (pre)	0.819	0.819	0.843	0.843	0.771	0.771
Observations	204,954	204,954	164,034	164,034	93,887	93,887
Author fixed effects	✓	✓	✓	✓	✓	✓
Cohort × Year fixed effects	✓	✓	✓	✓	✓	✓
Field × Year fixed effects	✓	✓	✓	✓	✓	✓

Note: This table gathers the FE OLS regressions for researchers working on sensitive topics. Columns (1)–(2) report estimates for the similarity of Hungarian-language publications with the researcher's own pre-2005 publications, columns (3)–(4) show the estimates for the similarity of English-language publications. Columns (5)–(6) for the similarity with the current research of government-allied researchers. Odd columns report the average post-2005 treatment effect and even columns report the decomposition of this effect by government mandate. All specifications include author, cohort × year and field × year fixed effects. *** p<0.01, ** p<0.05, * p<0.1. *** p<0.01, ** p<0.05, * p<0.1.

1.H — Global effect: Complementary Results

1. Migration

Table 1.H.1 : Effect on Migration, Hungary vs. Central Europe

	affiliation in other country for over 2 years		for over 3 years		emigrated	
	(1)	(2)	(3)	(4)	(5)	(6)
hungary × post = 1	0.0114***		0.0144***		0.0146***	
	(0.0038)		(0.0038)		(0.0046)	
hungary × post (1rst mandate)		0.0131***		0.0130***		0.0109***
		(0.0036)		(0.0035)		(0.0041)
hungary × post (2nd mandate)		0.0117**		0.0163***		0.0202***
		(0.0047)		(0.0047)		(0.0057)
hungary × post (3rd mandate)		0.0101*		0.0139***		0.0165**
		(0.0055)		(0.0053)		(0.0067)
Observations	203,938	203,938	203,938	203,938	203,938	203,938
R^2	0.667	0.667	0.683	0.683	0.806	0.806
Within R^2	0.000	0.000	0.000	0.000	0.000	0.001
Mean.Dep.Var.Pre	0.078	0.078	0.063	0.063	0.098	0.098
final_id fixed effects	✓	✓	✓	✓	✓	✓
year × entry cohort fixed effects	✓	✓	✓	✓	✓	✓
year × field fixed effects	✓	✓	✓	✓	✓	✓

Note: results are from Fixed-Effect OLS regression, for each outcome, on data from OpenAlex. The unit of observation is author by year and the sample period is from 2005-2020. Treated authors are authors from Hungary. Control authors are Slovenian researchers matched to the treated. The dependent variable is an indicator variable taking the value 1 if the author publishes with an affiliation outside of Hungary for the treated and Slovenia for the control after that date for two or three consecutive years (resp. columns (1), (2)), and another indicator variable taking the value 1 if the author is never seen publishing with an affiliation in Hungary for the treated and Slovenia for the control after that date (column (3)). Control variables account for trends for the matching cell over time. Standard errors (SE) are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.H.2 : Effect on International Migration - By Type of Affiliation

	Prob. of being only abroad	Prob. of both home and abroad	Prob. of being only abroad (4y)	Prob. of both home and abroad (4y)
	(1)	(2)	(3)	(4)
hungary × post (1rst mandate)	0.0340***	-0.0140	0.0240**	-0.0124*
	(0.0126)	(0.0121)	(0.0100)	(0.0074)
hungary × post (2nd mandate)	0.0202	-0.0163	0.0271**	-0.0085
	(0.0146)	(0.0135)	(0.0126)	(0.0089)
hungary × post (3rd mandate)	0.0296*	-0.0292**	0.0393***	-0.0133
	(0.0159)	(0.0138)	(0.0146)	(0.0089)
hungary × post (4th mandate)	0.0372*	-0.0509**	0.0341**	-0.0239**
	(0.0208)	(0.0210)	(0.0166)	(0.0111)
Observations	22,625	22,625	22,625	22,625
R^2	0.603	0.300	0.669	0.399
Within R^2	0.001	0.001	0.002	0.001
Pre-period mean	0.074	0.071	0.030	0.023
Author FE	✓	✓	✓	✓
Year-Field FE	✓	✓	✓	✓
Year-Entry year FE	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS. The unit of observation is author by year and the sample period is 2006–2023. Treated authors are from Hungary; control authors are Slovenian and Croatian researchers matched to the treated. Each column reports the decomposition of the treatment effect by political mandate. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.H.3 : Effect on International Migration - By Region of destination in Europe

	Prob. of being in Western Europe(4y) (1)	Prob. of being in Central Europe (4y - home country excluded) (2)	Prob. of being in Northern Europe (4y) (3)	Prob. of being in Eastern Europe (4y) (4)
hungary × post (1^{rst} mandate)	0.0035	-0.0016	0.0009	0.0003
	(0.0024)	(0.0018)	(0.0013)	(0.0005)
hungary × post (2nd mandate)	0.0064*	-0.0033	0.0001	0.0011
	(0.0033)	(0.0025)	(0.0016)	(0.0009)
hungary × post (3rd mandate)	0.0055	-0.0035	-0.0010	0.0017
	(0.0037)	(0.0028)	(0.0017)	(0.0014)
hungary × post (4th mandate)	0.0047	-0.0045	-0.0014	0.0021
	(0.0036)	(0.0031)	(0.0017)	(0.0014)
Observations	83,128	83,128	83,128	83,128
R^2	0.750	0.681	0.693	0.778
Within R^2	0.000	0.000	0.000	0.000
Pre-period mean	0.016	0.002	0.003	0.003
Author FE	✓	✓	✓	✓
Year-Field FE	✓	✓	✓	✓
Year-Entry year FE	✓	✓	✓	✓

Note: Results are from Fixed-Effect OLS regressions. The unit of observation is author by year and the sample period is 2006–2023. Treated authors are from Hungary; control authors are Slovenian and Croatian researchers matched to the treated. Each column reports the decomposition of the treatment effect by political mandate. Standard errors are clustered by author. *** p<0.01, ** p<0.05, * p<0.1.

1.I — Robustness Exercises

1. Within Hungary: Opponents and Allies

We run three falsification exercises on the within-Hungary design. All use the PPML specification of the main analysis (author, year × field and year × entry-year fixed effects, exact matching, standard errors clustered at the author level).

Placebo Event Study.

We re-estimate the event study on the pre-period only (1990–2010), so the real treatment never enters the sample, and assign a fake treatment date in 2006. Because 2006–2010 productivity lies inside this window, authors are matched on entry year and field only. For allies, almost all coefficients are insignificant, before and after the fake date (Figure 1.I.1). For opponents, the coefficients for publications and co-authors stay around -0.1 both before and after the fake date, with a few significant years on either side. This reflects a relatively high level in the 2005 reference year rather than a differential trend: there is no break at the fake date. Consistently, the difference-in-differences estimated for each fake treatment year between 2001 and 2007, which compares all years before and after the cutoff, is insignificant in 54 of the 56 estimates (4 outcomes, 7 years, 2 groups), about what chance alone would produce; the two exceptions are publications of opponents (fake year 2005) and citations of allies (fake year 2007) (Figure 1.I.2).

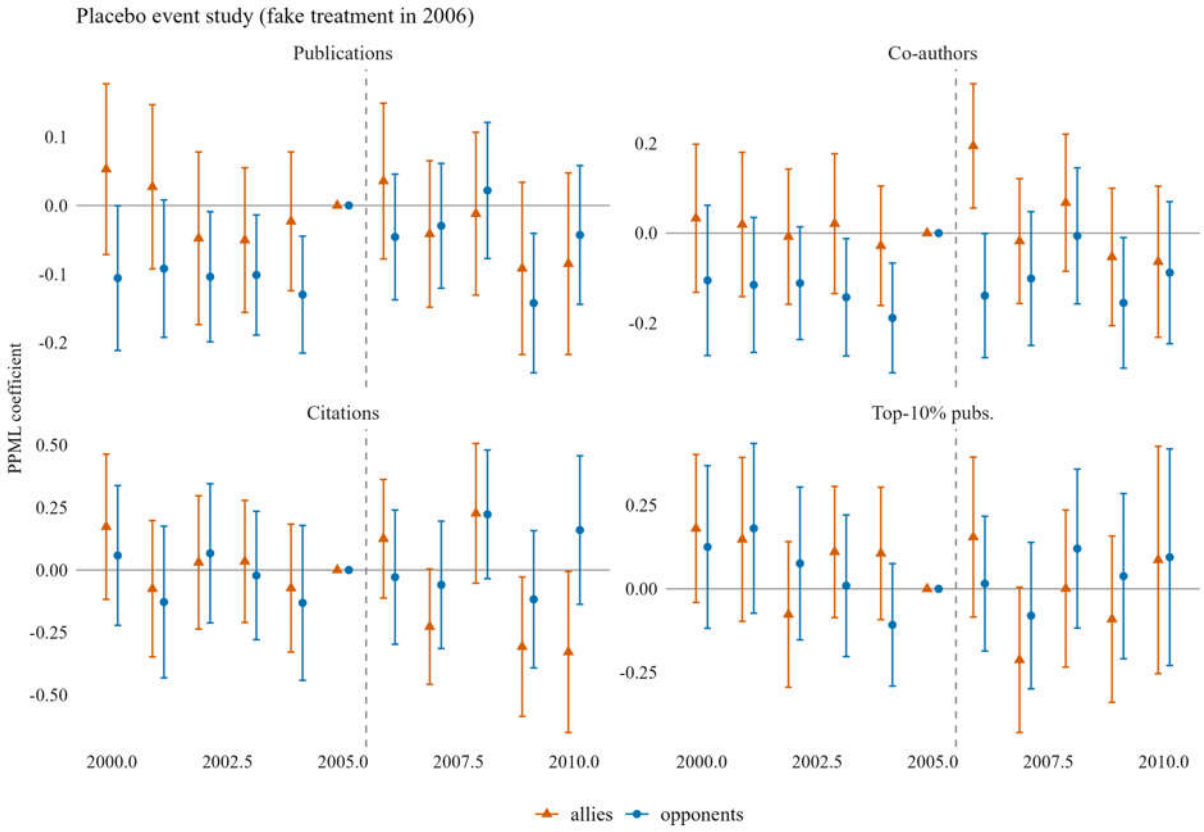
Random Treatment.

We draw 1,000 authors at random from the neutral researchers (actual opponents and allies are excluded), label them as treated, re-match and re-estimate. We repeat this 500 times. Under a correctly sized test, about 5% of draws should be significant at the 5% level; with 500 draws, rates between 3.1% and 6.9% are consistent with correct size. Figure 1.I.3 shows that the t-statistics are centred on zero, with rejection rates within this band.

Randomization Inference.

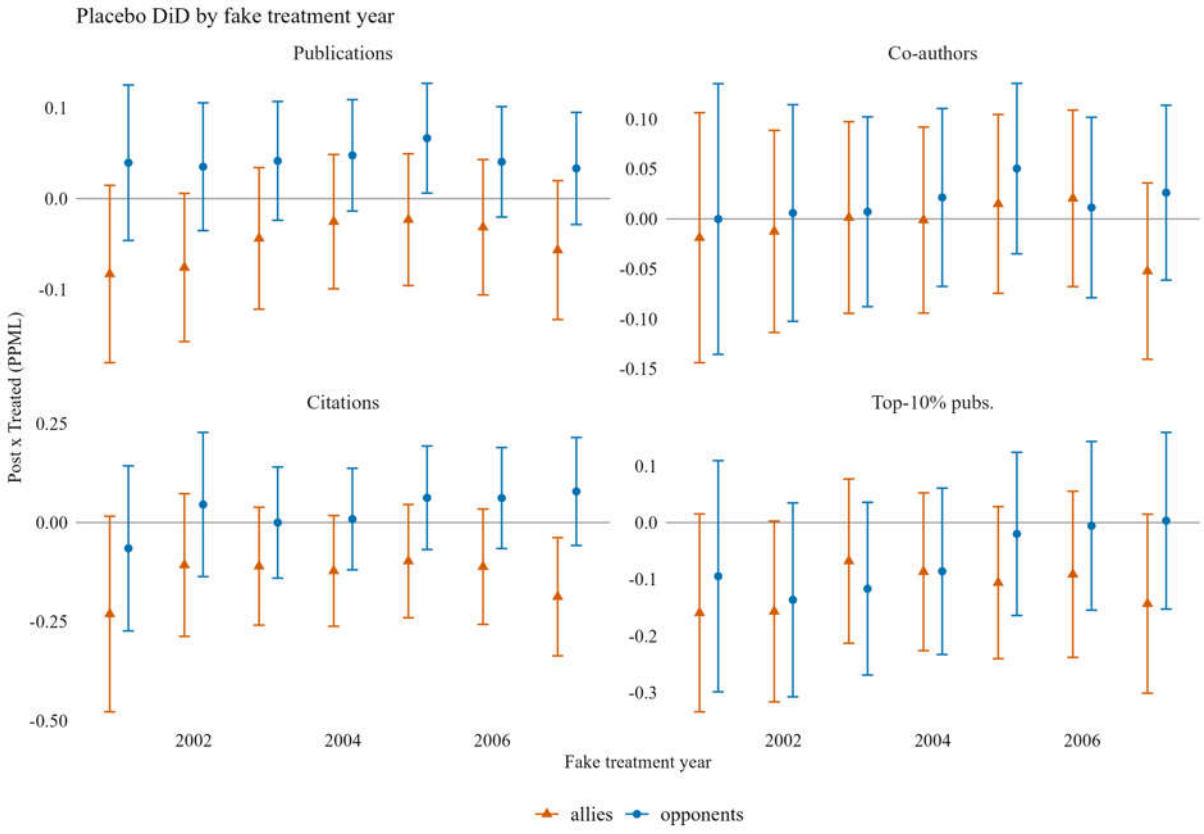
Finally, we permute the actual treatment labels across authors within matching strata 500 times and compare our main estimate with the resulting null distribution (Figure 1.I.4). For opponents, the randomization-inference p-values are close to the clustered ones: 0.477 (clustered: 0.478) for publications, 0.096 (0.087) for citations, 0.024 (0.025) for co-authors and 0.046 (0.012) for top-10% publications. The effects on co-authors and top-10% publications therefore remain significant at the 5% level under randomization inference.

Figure 1.I.1 : Placebo Event Study, Within Hungary



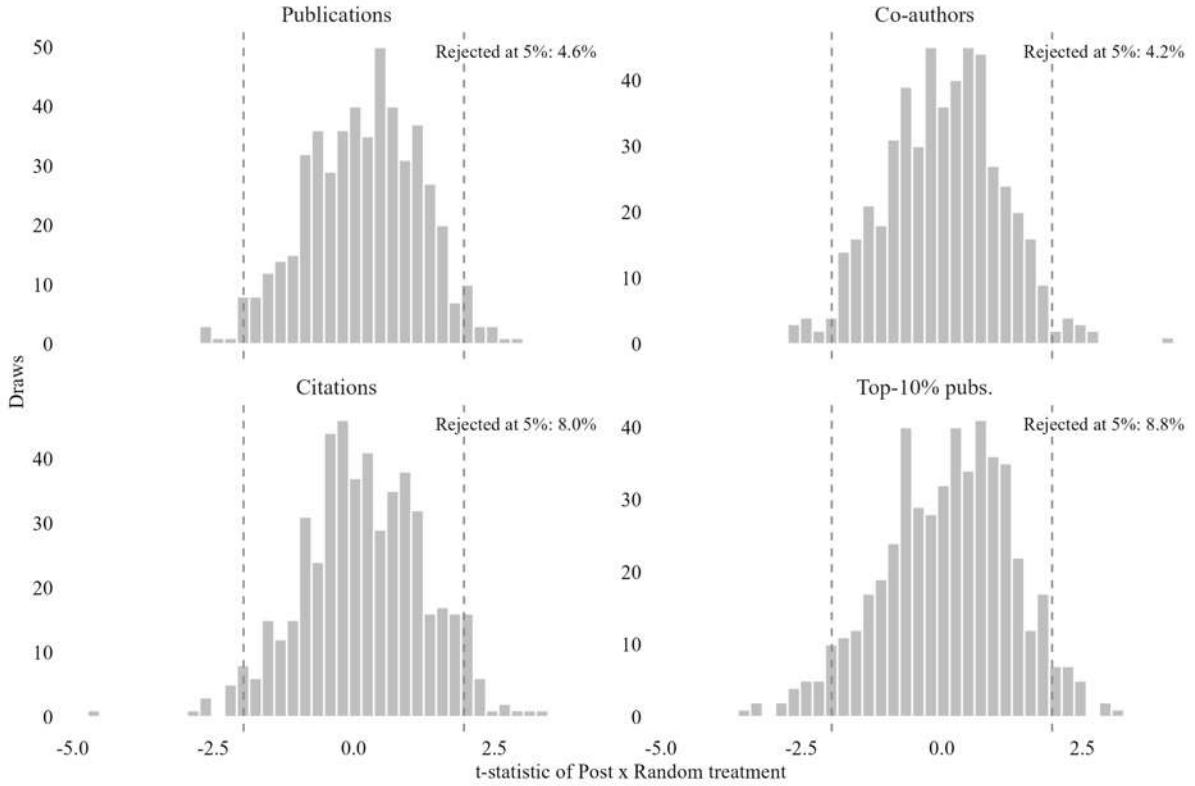
Notes: The figure reports year × treatment coefficients from Fixed-Effects PPML regressions using MTMT data over 1990–2010, with a fake treatment date in 2006 (reference year 2005). Blue: political opponents; orange: government allies; each group is compared with its matched control group of neutral researchers. Authors are exactly matched on entry year and field. Publications, co-authors and citations are winsorised at the top 1 percent. 95% confidence intervals use standard errors clustered at the author level.

Figure 1.I.2 : Placebo Difference-in-Differences by Fake Treatment Year, Within Hungary



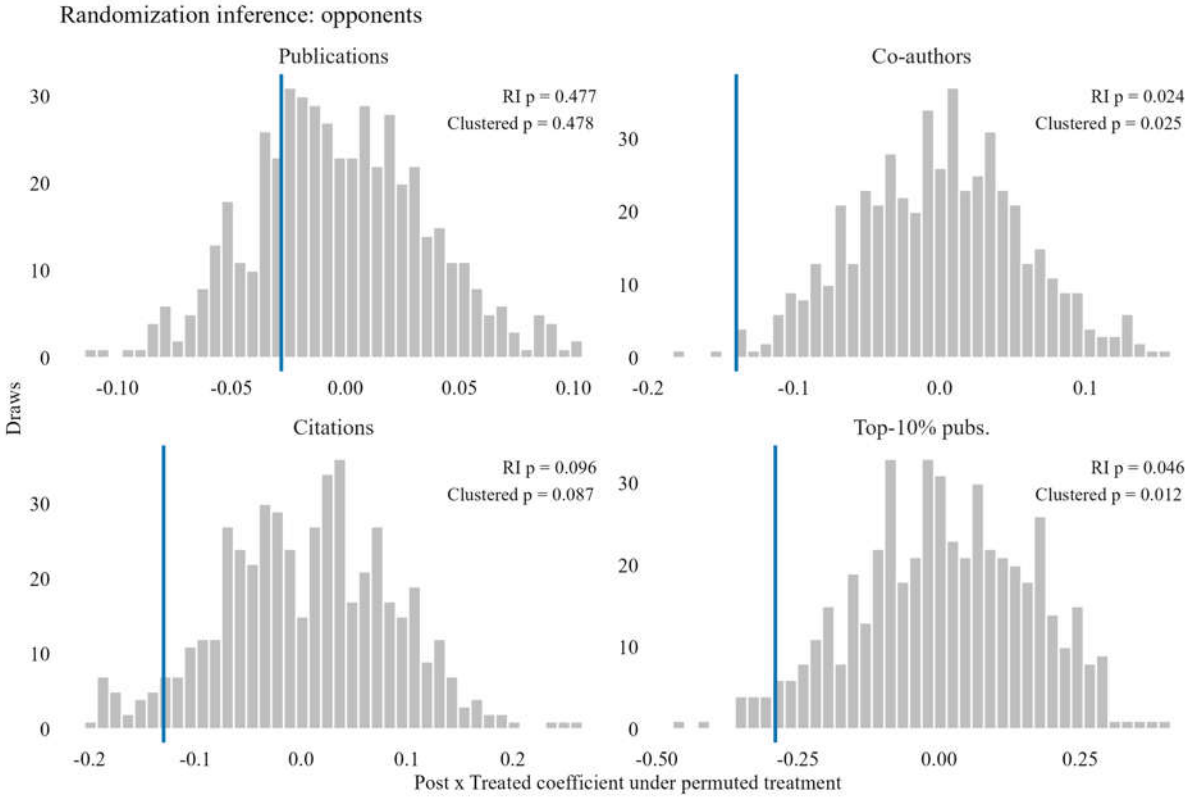
Notes: Each point is the Post × Treated coefficient from a separate Fixed-Effects PPML regression on MTMT data over 1990–2010, where Post equals one from the fake treatment year onwards. Blue: political opponents; orange: government allies. Matching, fixed effects and winsorization as in Figure 1.I.1. 95% confidence intervals use standard errors clustered at the author level.

Figure 1.I.3 : Random Treatment Assignment, Within Hungary



Notes: The figure shows the distribution of t-statistics of Post × Random treatment across 500 draws. In each draw, 1,000 authors are randomly drawn from researchers who are neither opponents nor allies and labelled as treated; the sample is then exactly matched on entry year, field and 2006–2010 productivity and the main Fixed-Effects PPML specification is re-estimated on MTMT data over 1990–2020 (Post: 2011 onwards). Dashed lines mark ±1.96. Standard errors are clustered at the author level.

Figure 1.I.4 : Randomization Inference, Political Opponents



Notes: Grey bars show the distribution of the Post × Treated coefficient when the actual treatment labels are permuted across authors within matching strata (500 permutations). The vertical line is the estimate with the actual labels. RI p is the share of permutations with an absolute t-statistic at least as large as the actual one; clustered p uses author-clustered standard errors. Specification as in the main analysis.

2. Cross-Country: Hungary vs. Slovenia and Croatia

With a single treated country, the main threat is a Hungary-wide shock unrelated to the political change, and inference based on country clusters is not informative. We therefore cluster standard errors at the country × field level and complement them with three exercises. All use a demanding version of the main specification: Fixed-Effects PPML on 2006–2023, Post from 2011, authors entering 1990–2005, coarsened exact matching on entry year, field, 2005–2010 publications, English-language publications, co-authors and top-10% publications, year × matching-stratum fixed effects, outcomes winsorised at the top 1 percent, and authors linked to the different political groups excluded.

Placebo Event Study.

We re-estimate the event study over 2000–2010 with a fake treatment in 2006, and the difference-in-differences for fake treatment years from 2003 to 2008, on authors publishing every year in 2000–2005 (Figure 1.I.5). Authors are matched on entry year, field, co-authors and top-10% publications measured over 2000–2005, so the fake post-period years (2006–2010) are the informative ones. For top-5% publications, no coefficient is significant, either in the event study or for any fake treatment year, although the estimates are imprecise. For co-authors, the fake post-period coefficients for 2006–2009 are not significant, and the fake-cutoff estimates are close to zero and insignificant for every year. Two deviations stand out: the 2000 and 2002 coefficients lie significantly below the 2005 reference, and the 2010 coefficient is negative and significant.

Random Treatment.

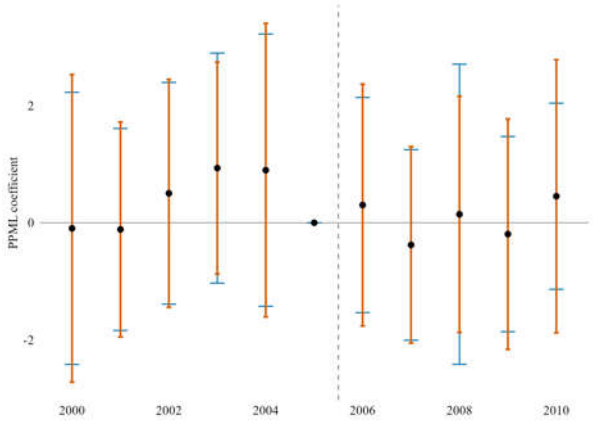
We randomly assign treatment to as many Slovenian and Croatian authors as there are matched Hungarian authors (Hungarian authors are excluded from the pool), re-match and re-estimate, 500 times. Under a correctly sized test, 5% of draws should be significant at the 5% level; with 500 draws, rates between 3.1% and 6.9% are consistent with correct size. For top-5% publications, for instance, 5.6% of draws are rejected with country × field clustering and 4.4% with author clustering (Figure 1.I.6).

Randomization Inference.

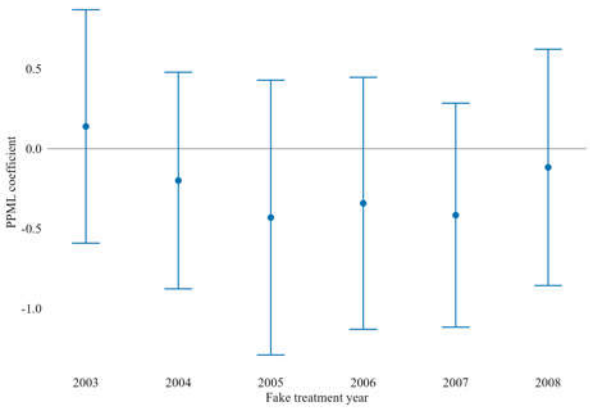
We permute the actual treatment labels across authors within matching strata 1,000 times and compare the main estimate with the resulting null distribution (Figure 1.I.7). The randomization-inference p-values are consistent with the clustered ones. Because treatment is assigned at the country level, this exercise does not account for country-wide shocks.

Figure 1.I.5 : Placebo Exercises, Cross-Country

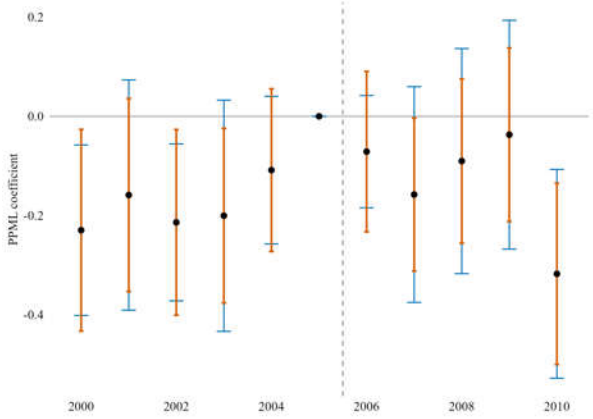
(a) Top-5% publications: event study



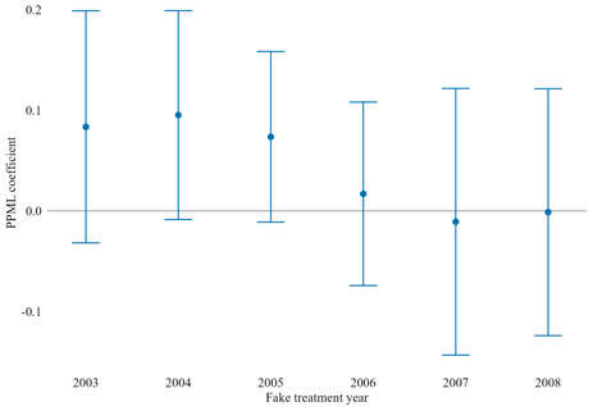
(b) Top-5% publications: fake treatment years



(c) Co-authors: event study

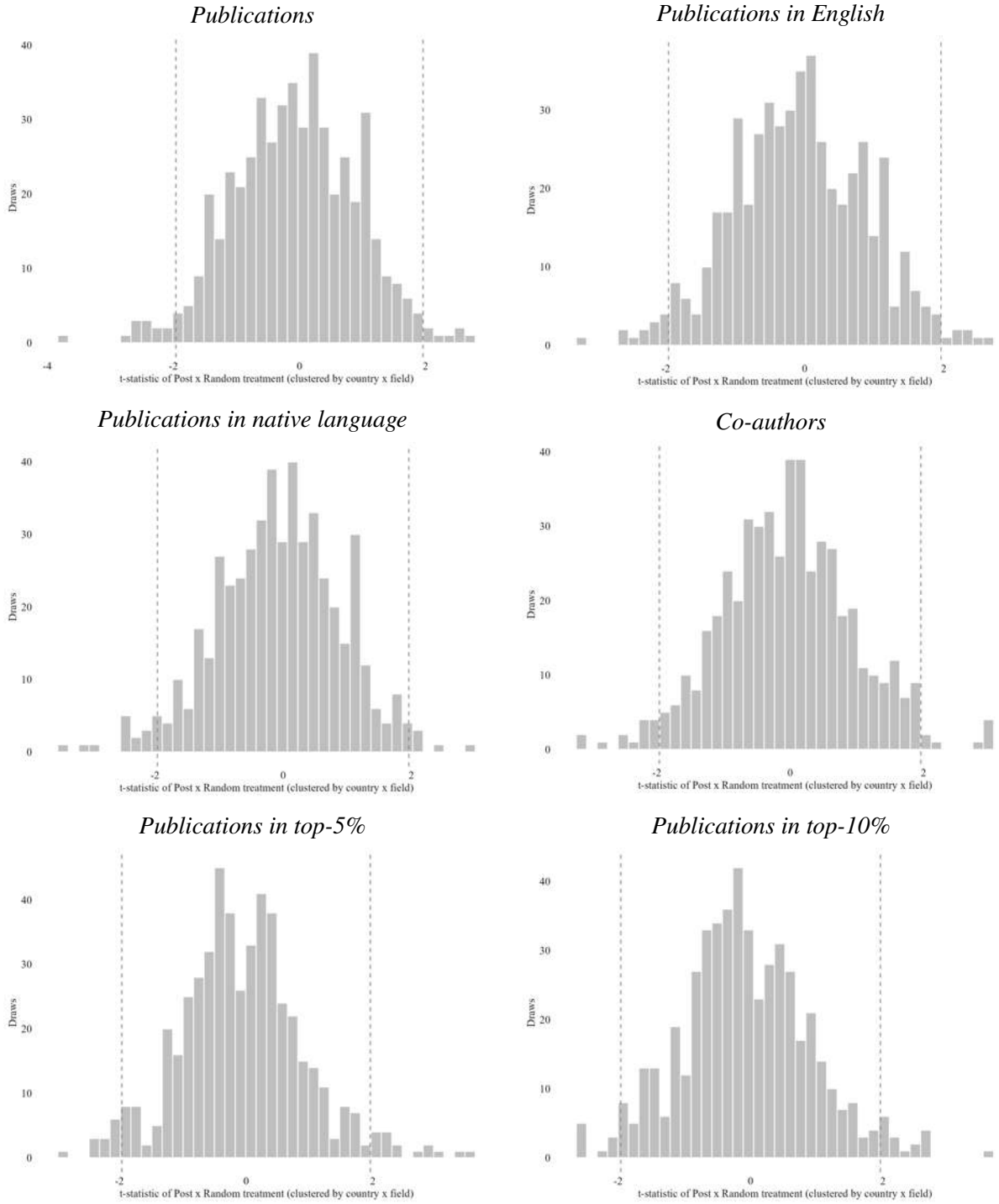


(d) Co-authors: fake treatment years



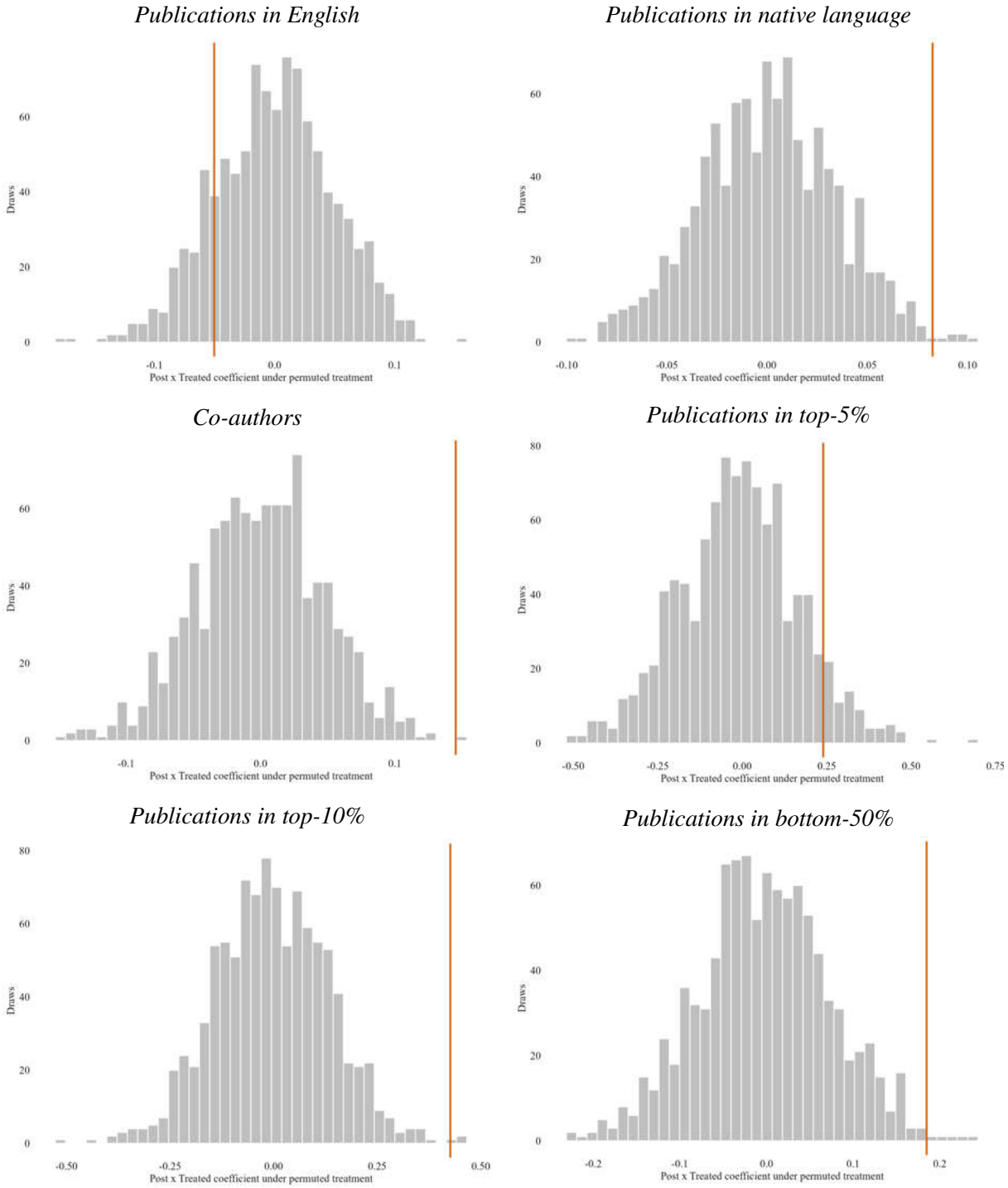
Notes: The figure reports placebo estimates from Fixed-Effects PPML regressions using data from CRIS national repositories over 2000–2010. Treated authors are researchers based in Hungary; controls are researchers from Slovenia and Croatia entering 1990–2000 and publishing every year in 2000–2005. Authors are matched on two-year entry-year bins, field, quintiles of 2000–2005 co-authors and bins of 2000–2005 top-10% publications; the regressions include author, year × field, year × entry year and year × matching-stratum fixed effects. Panels (a) and (c) report year × treatment coefficients with a fake treatment date in 2006 (reference year 2005); blue: 95% confidence intervals clustered by country × field ($t(G-1)$); orange: clustered by author. Panels (b) and (d) report the Post × Treated coefficient from separate regressions where Post equals one from the fake treatment year onwards, with 95% confidence intervals clustered by country × field. Outcomes are winsorised at the top 1 percent.

Figure 1.I.6 : Random Treatment Assignment, Cross-Country



Notes: Distribution of t-statistics of Post × Random treatment across 500 draws. In each draw, as many authors as there are matched Hungarian authors are randomly drawn from Slovenian and Croatian researchers and labelled as treated; the sample is re-matched and re-estimated. Dashed lines mark ±1.96. Standard errors clustered by country × field. Specification: Fixed-Effects PPML regressions using data from CRIS national repositories over 2006–2023, with Post equal to one from 2011. Treated authors are researchers based in Hungary; controls are researchers from Slovenia and Croatia entering 1990–2005. Authors are matched on two-year entry-year bins, field, deciles of 2005–2010 publications and English-language publications, quintiles of 2005–2010 co-authors and bins of 2005–2010 top-10% publications; the regressions include author, year × field, year × entry year and year × matching-stratum fixed effects. Outcomes are winsorised at the top 1 percent.

Figure 1.I.7 : Randomization Inference, Cross-Country



Notes: Grey bars show the distribution of the Post × Treated coefficient when the actual treatment labels are permuted across authors within matching strata (1,000 permutations). The vertical line is the estimate with the actual labels. Specification: Fixed-Effects PPML regressions using data from CRIS national repositories over 2006–2023, with Post equal to one from 2011. Treated authors are researchers based in Hungary; controls are researchers from Slovenia and Croatia entering 1990–2005. Authors are matched on two-year entry-year bins, field, deciles of 2005–2010 publications and English-language publications, quintiles of 2005–2010 co-authors and bins of 2005–2010 top-10% publications; the regressions include author, year × field, year × entry year and year × matching-stratum fixed effects. Outcomes are winsorised at the top 1 percent.

Annexe 2
Annexes du chapitre 2
Characterising the Gains to Autonomy in the Public Sector: Evidence from French Academia

2.A — Appendix to Section II

Through the architecture decided by the 1968 Faure law and the 1984 Savary law, French university governance was both very decentralised in appearance and highly centralised in practice. Indeed, while each university has an administrative council headed by a president, universities are divided into several components ("composantes", including labs, institutes and units devoted to both research and education, defined by field), each having a separate director, which diminishes the power of the president. The system thus operated at the conjunction of this de facto decentralised governance and the centralised financing, which entirely went through the government. The introduction of norms decided by the Lisbon Strategy allowed the government to attempt to change the organisation of universities. The goal was the following: by allowing universities more freedom to choose their research and education programs while incentivising them to be financially responsible through the transfer of management rights to their budget, the innovation potential of universities should be increased.

Funding of universities itself was reformed by the introduction of the "SYMPA" ("Système de répartition des Moyens à la Performance et à l'Activité", a performance- and activity-based allocation system) in which 20% of the budget was allocated based on university performance. However, Adnot (2019) contends that the system has mainly been used as a guide rather than an automatic way of allocating budget and that it has not managed to tie performance to funding in practice. The 2015 report by the French Court of Auditors (Cour des comptes, 2015), while praising the autonomy reform for its modernising effect, shows that it is incomplete and that most decisions regarding funding are not at the university level, and that there have been issues regarding adaptation of universities to their new position as managers of their workforce.

The law also permitted and actively encouraged the mergers of universities into larger, multidisciplinary universities. The rollout of the reform was dependent on voluntary request by universities – for instance, universities in Strasbourg merged as soon as 2009.[111] The main rationale for mergers was the lack of international visibility of French universities, which were relatively low on the Shanghai ranking in no small part due to their fragmentation across disciplines. Furthermore, proponents of the reform argued that merging universities would reduce the fixed administrative and organisational costs of university management. Some public institutions that already enjoyed some degree of autonomy were also included in the reform. 16 mergers were completed between 2009 and 2023, most occurring between 2016 and 2019.

[111] In some cases, universities that attempted to merge stayed separate, but in most cases they completed the merger.

2.B — Appendix to Section III

1. Sample selection

a) Institution-level Dataset

First, I keep all institutions active from 1980 to 2025 in OpenAlex. I then construct the sample by keeping French institutions as they are and aggregate the rest as "abroad"[112]. I also link those French institutions to their year of adoption of RCE and IDEX for those that received either or both, as well as their administrative status in French data. I then compute all affiliations each author is registered at every year. By taking the lag by year, I am able to derive where each researcher is coming from, be it a French institution or the notional "abroad" institution. Researchers in their first year of publication are recorded as coming from a notional "entry" institution. I then keep the lowest-level affiliation for each author in each set. For instance, if researcher A is recorded as affiliated both to a CNRS laboratory and CNRS itself, or to a university laboratory and the university itself, I only keep the laboratory. This greatly diminishes the number of double affiliations and allows me to compute the weights of $1/n$ for each researcher where n is the number of affiliations. I then attribute characteristics to researchers: their first year of publication, their first institution and country of publication, and I also derive their gender from their first name using the algorithm of Bérubé et al. (2020). This value is computed for many more researchers than in the sample below as even researchers with a single affiliation in France are included: 30,301,591 different researcher profiles are included in this analysis. I then derive researcher-institution-year level characteristics: length of the period during which the researcher has been observed at this institution and former place of affiliation. I then aggregate the total number of researchers per institution and domain per year, both weighted and unweighted, by author-level and researcher-institution-year level category. I then turn to publication outcomes. Likewise, publication outcomes are aggregated at the institution-domain-year level, with respect to the same categories. The panel is then sphericised, all variables taking the value 0 when the institution is not observed in that period, yielding 12,651 institution-domain pairs with at least one affiliated researcher for years 1980-2025. I keep institution-domain pairs that had active researchers prior to 2003, were not treated after 2012, did not belong to one of the three IDEX clusters that were cancelled, which had at least 5 active researchers on average in the years 2000-2002, and less than the last percentile of active researchers in that period for the sake of comparability. I also remove hospitals and companies from the sample as they are not purely higher education and research environments. This yields a final sample of 1,948 units. I provide descriptive statistics for these units in the following tables and figures, split by treatment status.

Year of treatment	N	Average # of researchers	Median # of researchers	Average # of foreign researchers	Median # of foreign researchers	Average # of publications	Median # of publications	Average # of citations	Median # of citations
0	2,800	16.3	8.4	4.4	1.7	26.4	15	1,155.9	459.5
2009	1,380	23	12.3	4.9	2.2	36	22	1,490.7	724.5
2010	2,820	22.6	12.9	5	2.4	39.7	23	1,575.5	748
2011	1,515	21.6	12	4.6	2.4	39.6	22	1,436	678
2012	1,225	21.6	11.6	4.1	2	36.7	20	1,364	557

Table 2.B.1 : Descriptive Statistics - Institution-level Outcomes pre-2007

This table summarises the outcome variables used in the analysis for the sample of institutions prior to 2007 by year of treatment. The number of researchers is computed by weighing each researcher by the inverse of her number of affiliations.

b) Researcher-level Dataset

I keep authors whose first publication occurred after 1965 and prior to 2003, whose first and last publications are separated by at least two years and who are at least active in France during 2 years and who have at least two publications in the country. I then compute individual-level characteristics, such as domain, field and subfield of study in the OpenAlex classification, or gender based on Bérubé et al. (2020). I then compute researcher-year level outcomes. A

[112] I consider that an institution in OpenAlex is French if its OpenAlex localisation is in France or its homepage url ends in ".fr".

Year of treatment	Life Sciences	Social Sciences	Physical Sciences	Health Sciences
0	138	75	189	158
2009	66	37	121	52
2010	151	93	204	116
2011	64	60	126	53
2012	59	57	82	47

Table 2.B.2 : Descriptive Statistics - Institutions by Domain

This table summarises the number of institutions per domain and year of treatment.

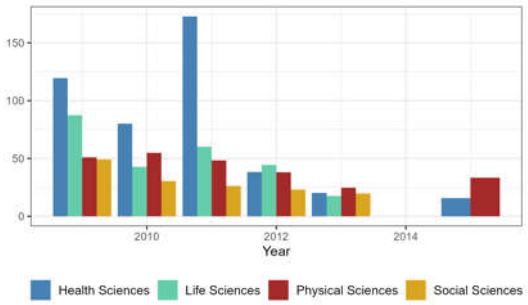


Figure 2.B.1 : Average Number of Authors in 2004-2007 by Domain of Research and Institution by Cohort of Treatment

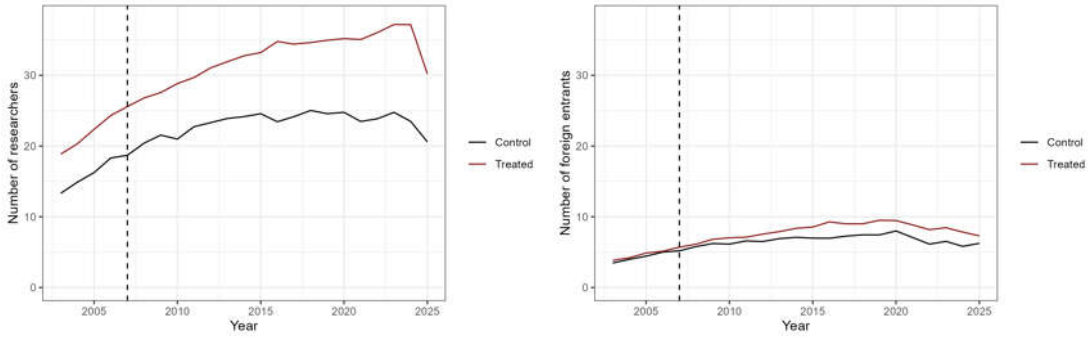
Notes: This graph depicts the average number of authors in each domain of science as identified by OpenAlex for each cohort of treatment for administrative autonomy.

researcher is considered to be affiliated to the same institution(s) until she publishes elsewhere. The panel is then sphericised for all years from 2003 to 2025, outcomes at 0 when not observed. I then remove authors whose field(s) contain fewer than 0.1% of the sample, and those who were ever affiliated to an institution with a cancelled IDEX, or a company or a hospital. I then keep authors who have at least one publication in the 2000-2002 period, i.e., who are active prior to the first year of the panel.

Year of treatment	University	Laboratory	Government	Non-profit	Archive and others	Share in CNRS	Share in Paris
0	77	223	83	39	138	0.152	0.309
2009	51	223	1	0	1	0.634	0.014
2010	105	447	9	0	3	0.578	0.138
2011	79	222	0	0	2	0.614	0.432
2012	36	209	0	0	0	0.690	0.131

Table 2.B.3 : Descriptive Statistics - Institution-level Characteristics

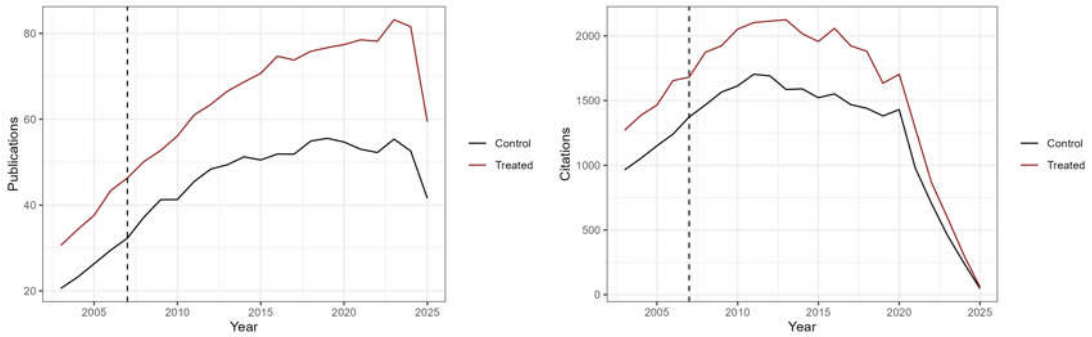
This table summarises the characteristics for the sample of institutions by year of treatment.



(a) Total (b) Foreign entrants

Figure 2.B.2 : Average Number of Researchers per Year in Treated and Control Institutions

Notes: This graph reports the average number of researchers (left) and number of researchers who first published outside of France (right) in treated and control institutions. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Publications (b) Citations

Figure 2.B.3 : Average Scientific Production per Year in Treated and Control Institutions

Notes: This graph reports the average number of publications (left) and citations (right) in treated and control institutions. Variables are winsorised at the 1% level.

Year of treatment	N	Average # of publications	Median # of publications	Average # of citations	Median # of citations	Average # of Top5% journal pub.	Average of novelty measure
0	24, 113	2.700	1	178.900	1	2	222.300
2009	6, 545	3.100	2	177.200	19	2.200	246.500
2010	17, 857	3.400	2	191.800	22	2.400	237
2011	10, 595	4.100	2	274	20	3.200	205.700
2012	7, 661	3.100	2	148.200	10	1.700	180.900

Table 2.B.4 : Descriptive Statistics - Researcher-Level Outcomes pre-2007

This table summarises the outcome variables used in the analysis for the sample of researchers prior to 2007 by year of treatment of their institution of affiliation in 2007.

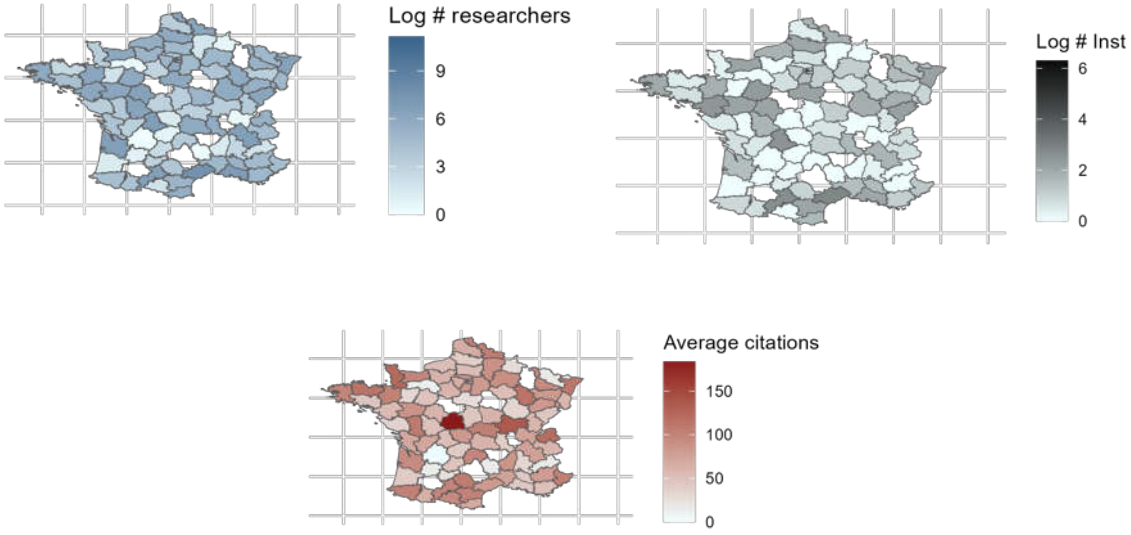


Figure 2.B.4 : Log of Number of Researchers (left), Number of Institutions (right) and Average Citations (bottom) by French Department

Notes: This graph reports the log number of authors, log number of higher education and research institutions and average citations per author active in each department in my regression sample. Citations are winsorised at the 1% level.

Year of treatment	Life Sciences	Social Sciences	Physical Sciences	Health Sciences
0	4,245	5,176	11,013	2,736
2009	1,232	1,216	3,516	367
2010	3,192	2,832	10,264	1,044
2011	1,225	2,535	6,185	401
2012	1,432	2,394	3,305	305

Table 2.B.5 : Descriptive Statistics - Researchers by Domain

This table summarises the number of researchers per domain and year of treatment.

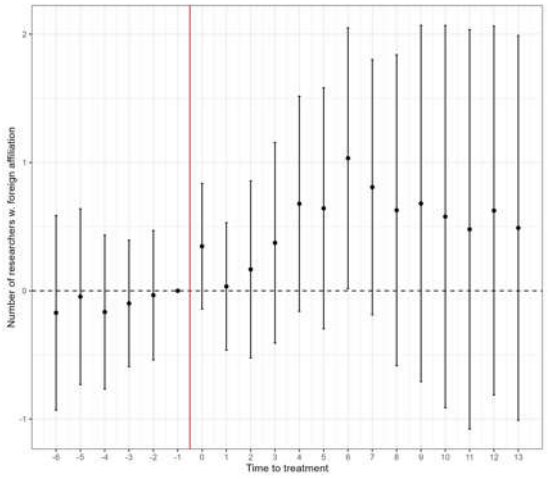


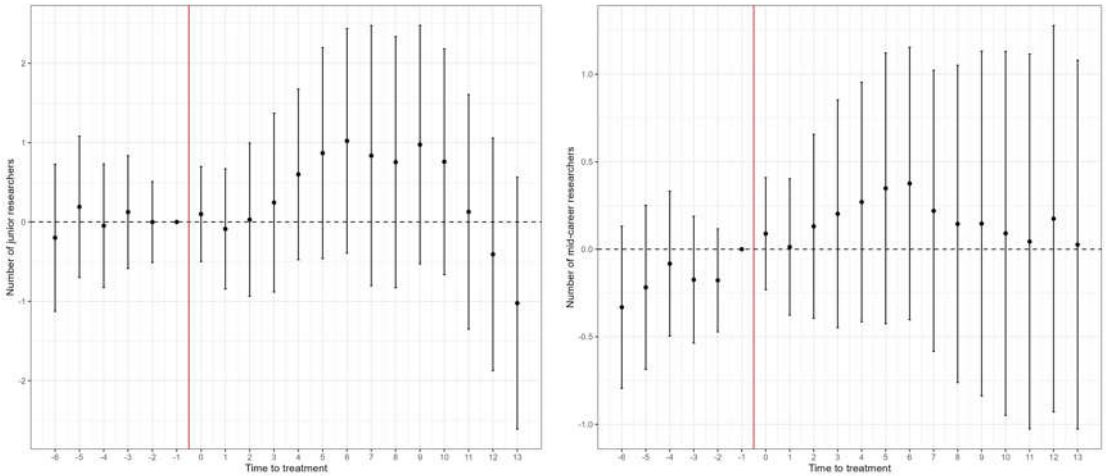
Figure 2.C.1 : Effect of Administrative Autonomy on Number of Researchers with Previous Affiliation Outside of France

Notes: This graph reports estimates of the effect on the total number of researchers whose last affiliation was outside of France per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

	Number of researchers (1)	*Number of foreign entrants* (2)	*Number of junior researchers* (3)	*Number of mid-career researchers* (4)	*Number of senior researchers* (5)	*Number of retiring researchers* (6)
All	7.302***	1.898***	0.488	1.980**	6.197***	1.542***
Std. error	(2.542)	(0.690)	(0.641)	(1.007)	(1.599)	(0.576)
Pre-treatment mean	42.712	8.934	14.428	9.531	11.457	3.468
N (units)	1,948	1,948	1,948	1,948	1,948	1,948
Controls						
Domain	Yes	Yes	Yes	Yes	Yes	Yes

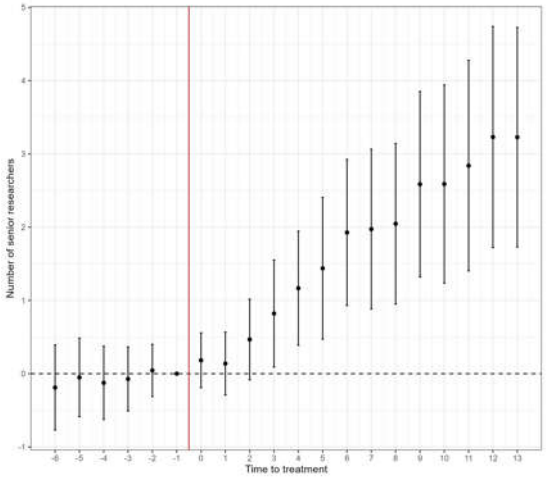
Table 2.C.1 : Effect of the Reform on Number of Researchers

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on the number of researchers in treated institutions per domain and per category: all (1), researchers who first published abroad (2), in the past 5 years (3), between 5 to 15 years prior to each year (4), and above 15 years (5), the number of researchers whose last publication was recorded this year (6) and the share of women (7). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as one researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.



(a) Junior

(b) Mid-career



(c) Senior

Figure 2.C.2 : Effect of Administrative Autonomy on Number of Researchers by Seniority

Notes: This graph reports estimates of the effect on the total number of researchers by seniority (under five years, between five and fifteen years, above fifteen years) per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

	Number of researchers (1)	*Number of foreign entrants* (2)	*Number of junior researchers* (3)	*Number of mid-career researchers* (4)	*Number of senior researchers* (5)	*Number of retiring researchers* (6)	*Share of women* (7)
All	5.526***	1.799***	1.991***	0.861***	2.064***	1.437***	0.005
Std. error	(1.156)	(0.553)	(0.493)	(0.274)	(0.343)	(0.291)	(0.008)
p-value	*0.000*	*0.001*	*0.000*	*0.002*	*0.000*	*0.000*	*0.532*
Pre-treatment mean	22.381	4.860	9.259	4.511	4.831	2.433	0.237
N (units)	1,948	1,948	1,948	1,948	1,948	1,948	1,745
Controls							
Domain	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.C.2 : Effect of the Reform on Number of Researchers - Data Truncated up to 2019

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on the number of researchers in treated institutions per domain and per category: all (1), researchers who first published abroad (2), in the past 5 years (3), between 5 to 15 years prior to each year (4), and above 15 years (5), the number of researchers whose last publication was recorded this year (6) and the share of women (7). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

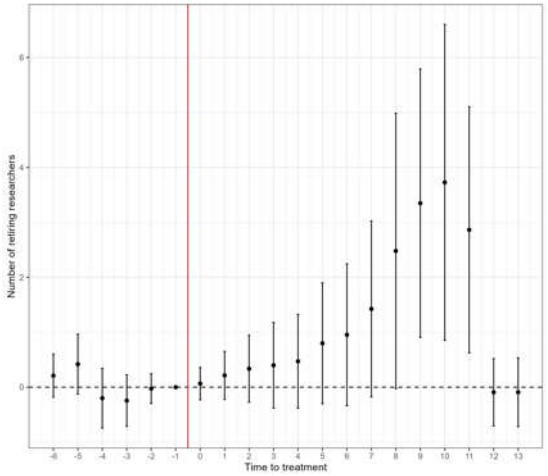


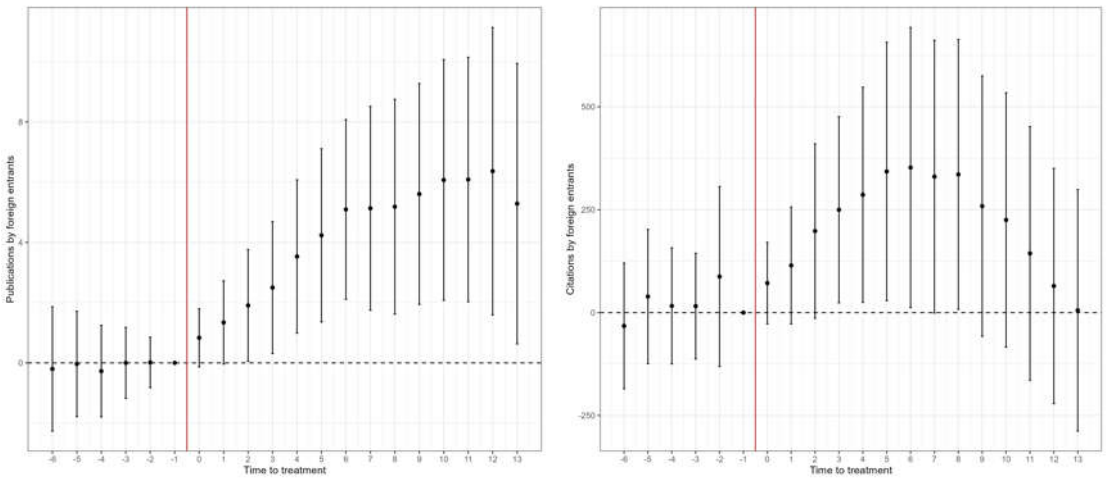
Figure 2.C.3 : Effect of Administrative Autonomy on Number of Researchers Retiring Each Year

Notes: This graph reports estimates of the effect on the total number of researchers retiring each year per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

	Number of stayers (1)	*Number of movers* (2)	*Number of foreign entrant stayers* (3)	*Number of foreign entrant movers* (4)
All	2.375***	1.064	0.589***	0.253
Std. error	(0.533)	(0.721)	(0.112)	(0.283)
Pre-treatment mean	8.316	13.922	1.117	3.610
N (units)	1,948	1,948	1,948	1,948
Controls				
Domain	Yes	Yes	Yes	Yes

Table 2.C.3 : Effect of the Reform on Number of Stayers and Movers

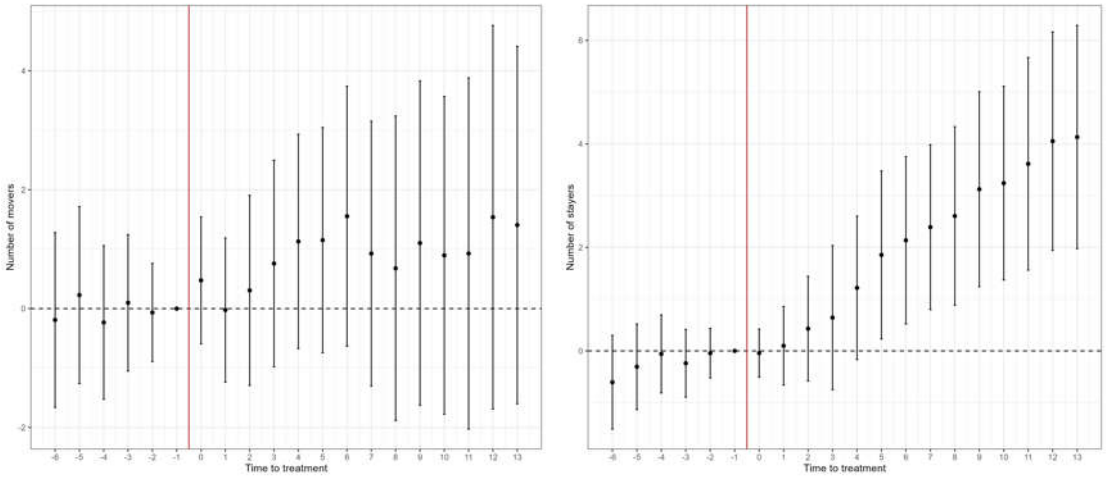
Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy for treated institutions on the number of researchers with last affiliation at this institution (1) and who just entered this institution (2), and on the same values for the subset of researchers who first published outside of France. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.



(a) Publications (b) Citations

Figure 2.C.4 : Effect of Administrative Autonomy on Total Production of Researchers who First Published Abroad

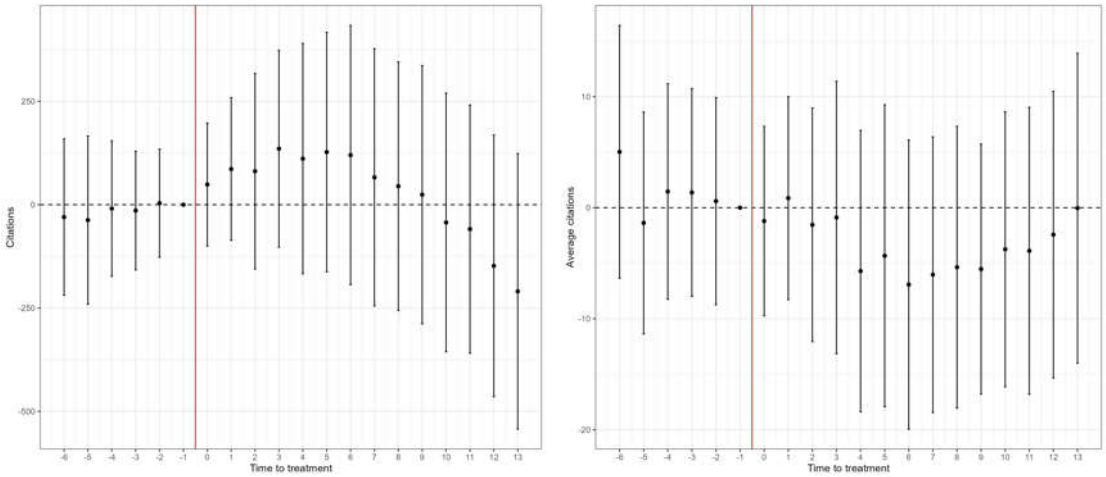
Notes: This graph reports estimates of the effect on the total publications (left) and citations (right) at this institution for researchers who first published abroad after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Movers (b) Stayers

Figure 2.C.5 : Effect of Administrative Autonomy on Number of Researchers by Previous Affiliation

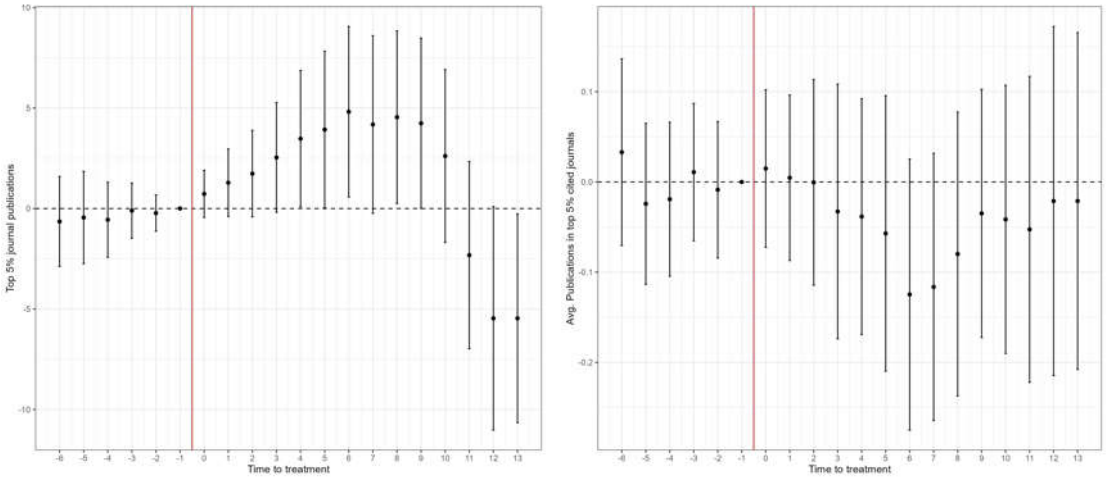
Notes: This graph reports estimates of the effect on the number of researchers arriving to the institution each year (left) and number of researchers who were affiliated to the institution the previous year (right) at this institution per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Total citations (b) Average Citations

Figure 2.C.6 : Effect of Administrative Autonomy on Citations of Researchers at Treated Institutions

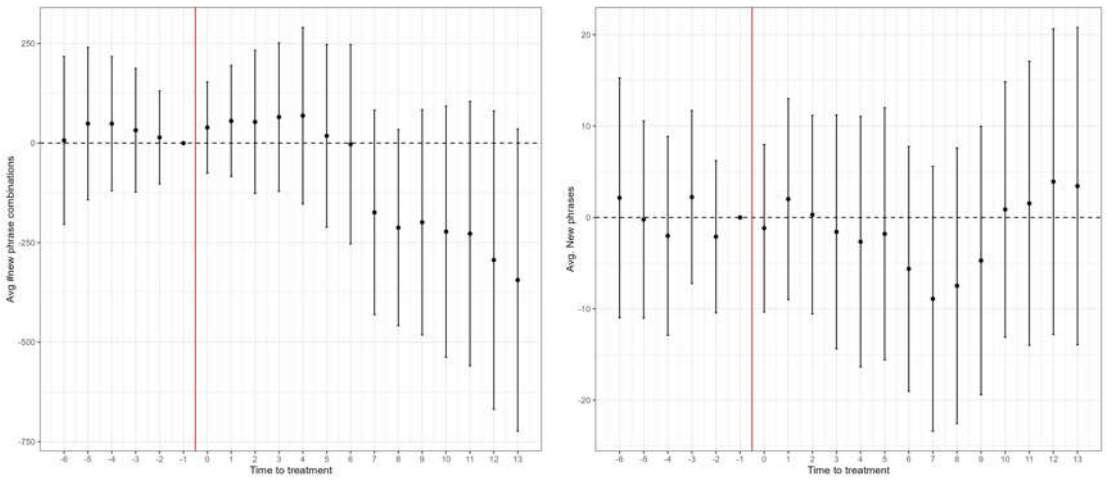
Notes: This graph reports estimates of the effect on the total citations (left) and average citations (right) of affiliated researchers for a treated institution per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Top Journal Publications (b) Average Top Journal Publications

Figure 2.C.7 : Effect of Administrative Autonomy on Publications of Treated Researchers in Top 5% Cited Journals

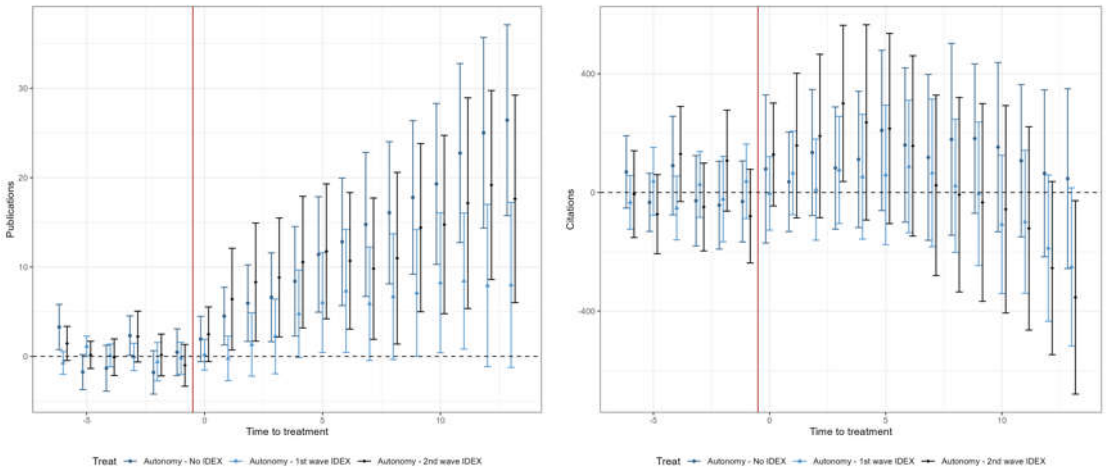
Notes: This graph reports estimates of the effect on the total publications (left) and average publications (right) in the top 5% most cited journals in OpenAlex of affiliated researchers for a treated institution per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Total (b) Average

Figure 2.C.8 : Effect of Administrative Autonomy on New Phrase Combinations Weighted by Later Reuse in OpenAlex

Notes: This graph reports estimates of the effect on the total (left) and average (right) number of new phrase combinations weighted by later reuse in OpenAlex (Arts, Melluso, and Veugelers, 2025) of affiliated researchers for a treated institution per year after treatment for administrative autonomy. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Publications (b) Citations

Figure 2.C.9 : Effect of Administrative Autonomy on Publications and Citations by IDEX wave of the Institution

Notes: This graph reports estimates of the effect on the publications (left) and citations (right) year after treatment for administrative autonomy. Estimates are obtained following Callaway and Sant'Anna (2021) in separate regressions for institutions that obtained an IDEX in the first wave (black), in the second wave (light blue) or never (dark blue). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

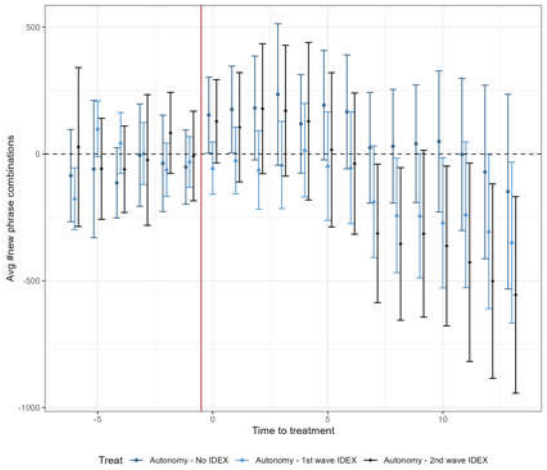
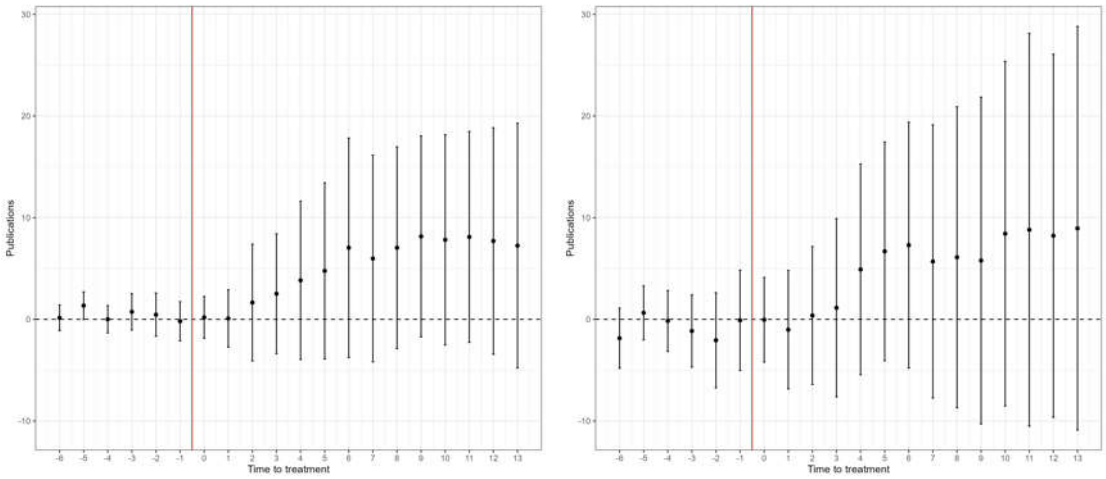


Figure 2.C.10 : Effect of Administrative Autonomy on New Phrase Combinations Weighted by their Reuse by IDEX Wave of the Institution

Notes: This graph reports estimates of the effect on the total number of new phrase combinations weighted by their reuse (Arts, Melluso, and Veugelers, 2025) year after treatment for administrative autonomy. Estimates are obtained following Callaway and Sant'Anna (2021) in separate regressions for institutions that obtained an IDEX in the first wave (black), in the second wave (light blue) or never (dark blue). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Below Median (b) Above Median

Figure 2.C.11 : Effect of Administrative Autonomy on Total Publications in First-Wave IDEX Universities, by 2000-2002 Productivity

Notes: This graph reports estimates of the effect on total publications in institutions that received an IDEX during the first wave of funding, for those institutions that counted a total number below the median by institution during the period 2000-2002 (left) and above it (right). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

	Number of researchers (1)	*Number of foreign entrants* (2)	*Number of researchers w. previous company affiliation* (3)	*Number of retiring researchers* (4)	*Share of women* (5)
Life Sciences	4.584***	1.541***	0.049**	0.620	0.014
Std. error	(1.580)	(0.546)	(0.020)	(0.455)	(0.019)
Pre-treatment mean	16.785	4.371	0.106	1.586	0.284
N (units)	478	478	478	478	478
Social Sciences	10.278***	0.482	0.060*	1.726	-0.014
Std. error	(3.196)	(1.247)	(0.036)	(1.083)	(0.016)
Pre-treatment mean	21.623	3.462	0.063	2.522	0.312
N (units)	322	322	322	322	322
Physical Sciences	7.036***	3.069**	0.048	0.975*	0.002
Std. error	(2.399)	(1.254)	(0.052)	(0.566)	(0.009)
Pre-treatment mean	29.864	6.247	0.290	3.210	0.146
N (units)	722	722	722	722	722
Health Sciences	7.632*	-0.308	0.049*	2.121**	0.004
Std. error	(4.497)	(2.800)	(0.029)	(1.049)	(0.021)
Pre-treatment mean	15.298	4.009	0.112	1.878	0.256
N (units)	426	426	426	426	426
Controls					
Institution Type	Yes	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes

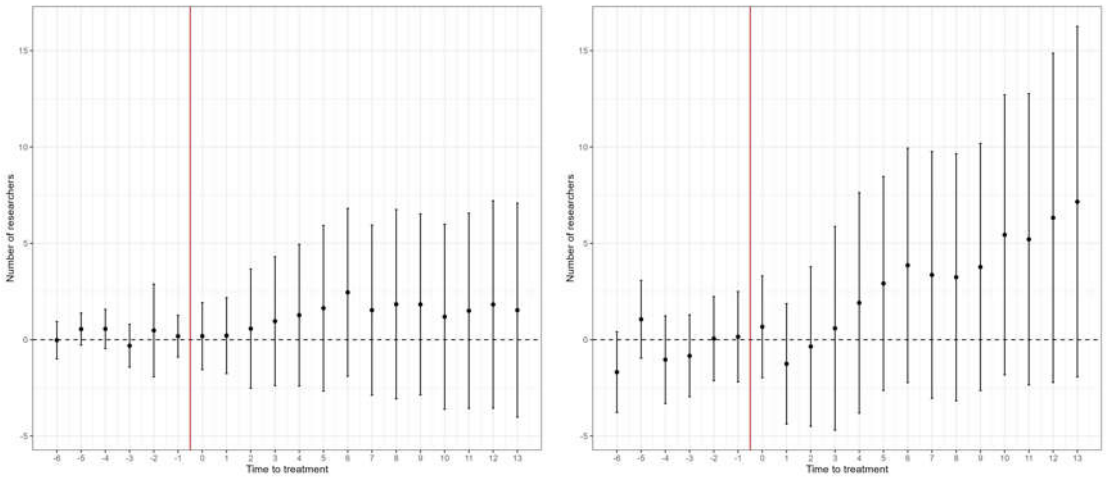
Table 2.C.4 : Effect of the Reform on Number of Researchers by Domain of Institution

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy per domain on the number of researchers in treated institutions per domain and per category: all (1), researchers who first published abroad (2), in the past 5 years (3), between 5 to 15 years prior to each year (4), and above 15 years (5), the number of researchers whose last publication was recorded this year (6) and the share of women (7). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Publications (1)	*Average publications* (2)	*Citations* (3)	*Average citations* (4)	*Top 5% journal publications* (5)	*Avg. Publications in top 5% cited journals* (6)	*Avg #new phrase combinations* (7)	*Avg. New phrases* (8)
Life Sciences	9.875**	0.363	209.013	32.114***	3.750**	0.123	-90.170	29.720
Std. error	(4.157)	(0.306)	(163.616)	(12.026)	(1.678)	(0.162)	(255.992)	(19.711)
Pre-treatment mean	25.528	1.931	1593.679	107.522	15.951	1.144	2735.535	183.243
N (units)	478	478	478	478	478	478	478	478
Social Sciences	14.706**	-0.611***	-126.586	0.829	-1.427	-0.135	-161.780***	-15.017**
Std. error	(7.228)	(0.172)	(180.353)	(29.694)	(0.913)	(0.129)	(50.517)	(6.643)
Pre-treatment mean	45.792	2.414	596.931	51.252	4.761	0.395	288.346	30.693
N (units)	322	322	322	322	322	322	322	322
Physical Sciences	12.526	-0.264*	-70.102	5.221	1.136	-0.024	-93.254	21.892
Std. error	(7.995)	(0.156)	(239.712)	(10.536)	(3.910)	(0.118)	(204.771)	(18.926)
Pre-treatment mean	52.823	2.066	2017.529	88.534	26.961	1.082	2446.891	110.146
N (units)	722	722	722	722	722	722	722	722
Health Sciences	24.747***	-0.713*	101.726	-17.553	4.885**	-0.295*	-563.866	-21.313
Std. error	(6.521)	(0.406)	(135.582)	(14.940)	(1.990)	(0.162)	(448.100)	(25.488)
Pre-treatment mean	24.383	2.269	1217.031	116.958	12.206	1.113	2344.469	192.075
N (units)	426	426	426	426	426	426	426	426
Controls								
Institution Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.C.5 : Effect of the Reform on Productivity of Affiliated Researchers by Domain of Institution

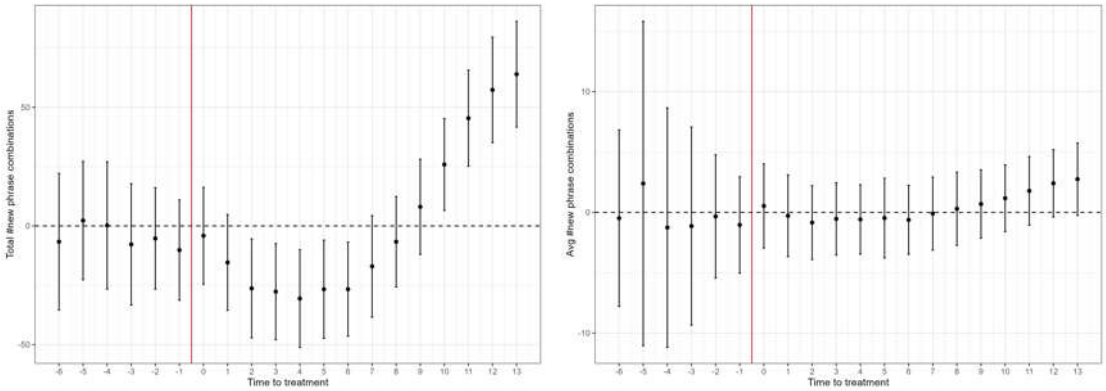
Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy per domain for treated institutions on total publications (1), average publications (2), total citations (3), average citations (4), total publications in the 5% most cited journals (5), average publications in the top 5% most cited journals (6), number of new phrase combinations weighted by reuse (7) and average number of new phrase combinations weighted by reuse (8) of researchers affiliated with the institution. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.



(a) Below Median (b) Above Median

Figure 2.C.12 : Effect of Administrative Autonomy on Number of Researchers in First-Wave IDEX Universities, by 2000-2002 Productivity

Notes: This graph reports estimates of the effect on the number of researchers in institutions that received an IDEX during the first wave of funding, for those institutions that counted a total number below the median by institution during the period 2000-2002 (left) and above it (right). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific domain, CNRS status, type of institution and quartile of institution in terms of total publications between 2000 and 2002. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.



(a) Total (b) Average

Figure 2.D.1 : Effect of Administrative Autonomy on Novelty of Research

Notes: This graph reports estimates of the effect on the total number per year and average number per paper of new phrase combinations weighted by their later reuse in OpenAlex for researchers who were affiliated to a treated institution in 2007. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific field, five year bin of year of first publication, quartile in terms of total publications and total citations between 2000 and 2002. Variables are winsorised at the 1% level.

	Publications (1)	*Top 5% journal publications* (2)	*Citations* (3)	*Total #new phrase combinations* (4)	*Avg #new phrase combinations* (5)
All	-0.777***	-0.737***	-47.591***	-20.942***	-0.327
Std. error	(0.069)	(0.056)	(5.733)	(5.954)	(0.919)
Pre-treatment mean	3.473	2.415	202.140	220.596	44.296
N (units)	66,771	66,771	66,771	66,771	66,771
Controls					
Entry year (5y. bins)	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes
Quartile of Cit. (selection)	Yes	Yes	Yes	Yes	Yes

Table 2.D.1 : Effect of the Reform on Productivity at the Individual Level - Limiting the Period to 2019

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on total publications (1), publications in the top 5% most cited journals (2), total citations (3), total (4) and average (5) number of new phrase combinations weighted by their reuse on researchers whose affiliation in 2007 was subsequently treated. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Prob. of publishing (1)	*Publications* (2)	*Top 5% journal publications* (3)	*Citations* (4)	*Total #new phrase combinations* (5)	*Avg #new phrase combinations* (6)
All	-0.046***	-1.087***	-0.334***	-31.194***	6.294	0.715
Std. error	(0.005)	(0.077)	(0.054)	(5.572)	(6.447)	(0.970)
Pre-treatment mean	0.767	3.477	2.417	202.362	220.839	44.345
N (units)	66,771	66,771	66,771	66,771	66,771	66,771
Controls						
Entry year (5y. bins)	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Cit. (selection)	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.D.2 : Effect of the Reform on Productivity at the Individual Level Conditional on not Having Left Academia

This table presents the Average Treatment Effect on the Treated of administrative autonomy on the probability of publishing (1), number of publications (2), publications in the 5% most cited journals (3), citations (4), and total number (5) and average number (6) of new phrase combinations weighted by their reuse, for researchers whose affiliation in 2007 was treated, conditional on the researcher publishing at least once in years after the year of observation. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. Standard errors are clustered at the 2007 affiliation level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Prob. of publishing (1)	*Publications* (2)	*Top 5% journal publications* (3)	*Citations* (4)	*Total #new phrase combinations* (5)	*Avg #new phrase combinations* (6)
No exit at all	-0.048***	-1.266***	-0.538***	-49.535***	-2.121	0.321
Std. error	(0.005)	(0.100)	(0.069)	(8.654)	(8.607)	(1.172)
Pre-treatment mean	0.820	4.377	3.180	277.378	287.266	51.342
N (units)	31,740	31,740	31,740	31,740	31,740	31,740
Controls						
Entry year (5y. bins)	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Cit. (selection)	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.D.3 : Effect of the Reform on Productivity at the Individual Level Conditional on not Having Left Academia in 2025

This table presents the Average Treatment Effect on the Treated of administrative autonomy on the probability of publishing (1), number of publications (2), publications in the 5% most cited journals (3), citations (4), and total number (5) and average number (6) of new phrase combinations weighted by their reuse, for researchers whose affiliation in 2007 was treated, conditional on the researcher being observed publishing in OpenAlex in 2025. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. Standard errors are clustered at the 2007 affiliation level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

2.E — Appendix to Section VI

1. Varying the Event Study Designs

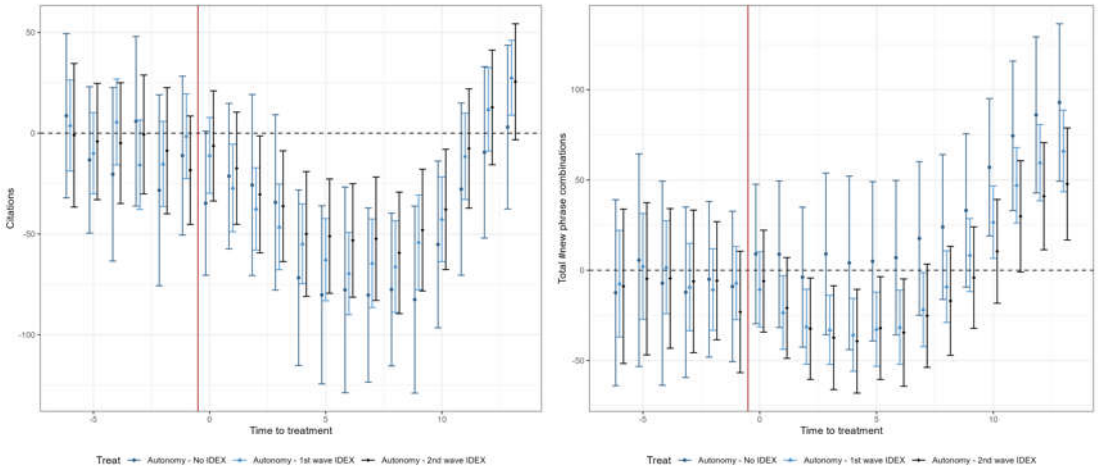
In this subsection, I vary the event-study designs in order to assess robustness of the results to the set of control researchers and explanatory variables.

a) Without Controls

I first provide the results of an estimation of Equation 2.1 on the first sample at the institution-level without any control to show that results are robust to this specification. In other words, the estimation is no longer doubly robust, but simply staggered.

Figure 2.E.1 shows the event study estimates for two of the main variables of the first set of results, total number of researchers (inversely weighted by number of affiliations) and total number of publications. Both trajectories strongly resemble the effects shown in Section IV, with similar magnitudes. Table 2.E.1 and Table 2.E.2 provide the results for the estimations presented respectively in Table 2.1 and Table 2.3 when removing all the controls from the estimation.

Similarly, Table 2.E.3 represents the same ATT as Table 2.4, when no controls are included. Results still appear to hold when trends are compared unconditionally.



(a) Publications (b) Citations

Figure 2.D.2 : Effect of Administrative Autonomy on Publications and Novelty by IDEX Wave of the Institution

Notes: This graph reports estimates of the effect on the publications and novelty proxied by total new phrase combinations weighted by their later reuse in OpenAlex for researchers who were affiliated to a treated institution in 2007. Estimates are obtained following Callaway and Sant'Anna (2021) in separate regressions for institutions that obtained an IDEX in the first wave (black), in the second wave (light blue) or never (dark blue). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific field, five year bin of year of first publication, quartile in terms of total publications and total citations between 2000 and 2002. Variables are winsorised at the 1% level.

	Number of researchers (1)	*Number of foreign entrants* (2)	*Number of junior researchers* (3)	*Number of mid-career researchers* (4)	*Number of senior researchers* (5)	*Number of retiring researchers* (6)	*Share of women* (7)
All	3.599***	0.970***	0.064	0.173	2.050***	0.630***	-0.001
Std. error	(1.010)	(0.326)	(0.359)	(0.233)	(0.349)	(0.190)	(0.006)
Pre-treatment mean	22.290	4.746	9.095	4.507	4.829	2.315	0.230
N (units)	1,948	1,948	1,948	1,948	1,948	1,948	1,948

Table 2.E.1 : Effect of the Reform on Number of Researchers - No Control Variables

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on the number of researchers in treated institutions per domain and per category: all (1), researchers who first published abroad (2), in the past 5 years (3), between 5 to 15 years prior to each year (4), and above 15 years (5), the number of researchers whose last publication was recorded this year (6) and the share of women (7). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

b) Using Only the Never-Treated as a Comparison

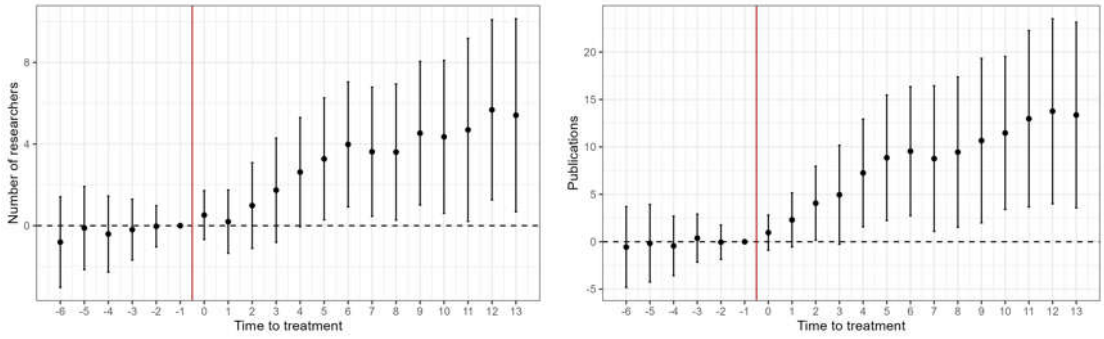
I now provide the results of an estimation of Equation 2.1 on each sample using the same set of controls, but comparing treated units to never-treated units only.

Again, Table 2.E.4 and Table 2.E.5 serve as the counterparts of Table 2.1 and Table 2.3 of the main text. While point estimates differ, although still in the same confidence interval, results appear mostly unchanged.

Table 2.E.6 presents the ATT on researchers based on their 2007 affiliation when only compared to the never-treated. The average estimate does not differ significantly from those presented in the main text in Table 2.4.

c) Estimates by Groups from Core Regressions

Figure 2.E.2 presents the individual effect of the reform by cohort of treatment adoption.



(a) Number of Researchers (b) Publications

Figure 2.E.1 : Effect of Administrative Autonomy on Total Researchers and Publications - No Control Variables

Notes: This graph reports estimates of the effect on total number of researchers (left) and publications (right) in institutions that received an IDEX. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level.

	Publications (1)	*Average publications* (2)	*Citations* (3)	*Average citations* (4)	*Top 5% journal publications* (5)	*Avg. Publications in top 5% cited journals* (6)	*Avg #new phrase combinations* (7)	*Avg. New phrases* (8)
All	9.030***	-0.019	-23.284	0.539	0.611	-0.023	-107.296	1.217
Std. error	(2.295)	(0.067)	(72.404)	(3.725)	(0.810)	(0.039)	(83.464)	(5.118)
Pre-treatment mean	38.432	2.064	1490.831	82.925	17.464	0.939	2066.979	117.670
N (units)	1,948	1,948	1,948	1,948	1,948	1,948	1,948	1,948

Table 2.E.2 : Effect of the Reform on Productivity at the Institution-Level - No Control Variables

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy for treated institutions on total publications (1) and total citations (2) by researchers who became affiliated with the lab in the past 5 years, and total publications (3) and total citations (4) by researchers affiliated with the lab who first published abroad. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

2. Varying the Year in which Treatment is Assigned

In this subsection, I present the overall estimates obtained when instead of considering that an author is treated in the same year as her 2007 affiliation, I consider that she is treated in 2006, 2008 or 2009. The estimation process is strictly the same as 2007, using the same control variables based on the period 2000-2002. Standard-errors are clustered at the level of researcher affiliation at the time of treatment assignment (i.e., in 2007 in the main text and 2006, 2008 and 2009).

Figure 2.E.3 shows the results of this exercise. First, it appears that all choices of years yield significant results of the same sign. Second, it appears that choosing 2008 or 2009 as the year to assign treatment yields more negative results. There is however no proof that this increase in effect is really causal, as researchers were able to make decisions to move or change their activities based on when universities would be granted autonomy as soon as 2007. Given that the 2007 estimates are not statistically significantly different from the more conservative 2006 estimates and are based on affiliation at the time the reform was passed, they are the favoured estimates for this analysis.

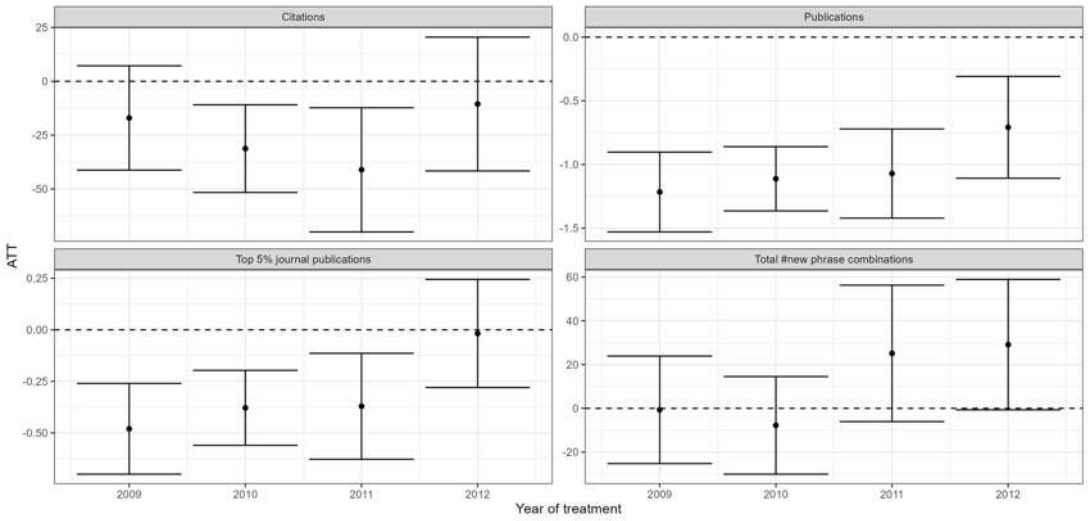


Figure 2.E.2 : ATT by Cohort of Adoption

Notes: This figure presents the Average Treatment Effect on the Treated of administrative autonomy for researchers in treated institutions, by cohort of adoption, on citations, publications, publications in the 5% most cited journals, and total number of new phrase combinations weighted by their later reuse in OpenAlex. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated units as controls. The model controls for scientific field, five year bin of year of first publication, quartile in terms of total publications and total citations between 2000 and 2002. Variables are winsorised at the 1% level.

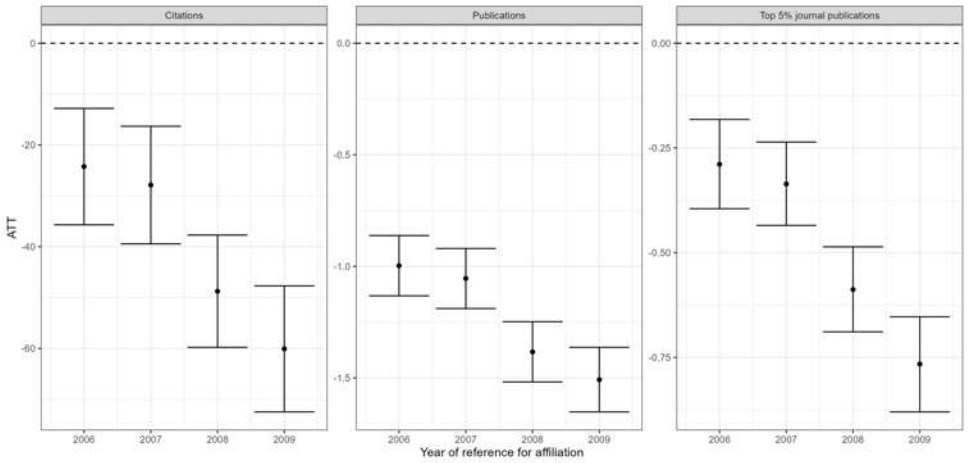


Figure 2.E.3 : Effect on Researchers Based on Year of Treatment of Affiliation, Varying the Year in which Treatment is Assigned

Notes: This graph reports estimates of the ATT on the citations (left), publications (center) and top 5% cited journals (right) per year after treatment for researchers affiliated to treated institutions, depending on the year chosen to fix treatment. The left-most point of each graph thus represents the estimate for treatment when considering that a researcher affiliated to a treated institution in 2006 is treated at the time her institution is treated, and the right-most point the estimate for treatment when considering that a researcher affiliated to a treated institution in 2009 is treated at the time her institution is treated. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. The model controls for scientific field, five-year bin of year of first publication, quartile in terms of total publications and total citations between 2000 and 2002. Variables are winsorised at the 1% level.

	Publications (1)	*Top 5% journal publications* (2)	*Citations* (3)	*Total #new phrase combinations* (4)	*Avg #new phrase combinations* (5)
All	-1.214***	-0.536***	-42.109***	-2.609	0.135
Std. error	(0.069)	(0.053)	(5.879)	(6.190)	(1.101)
Pre-treatment mean	3.473	2.415	202.140	220.596	44.296
N (units)	66,771	66,771	66,771	66,771	66,771

Table 2.E.3 : Effect of the Reform on Productivity at the Individual Level - No Control Variables

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on total publications (1), total citations (2), total publications in the top 5% most cited journals (3), total (4) and average (5) number of new phrase combinations weighted by their reuse on researchers whose affiliation in 2007 was subsequently treated. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated and not-yet-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Number of researchers (1)	*Number of foreign entrants* (2)	*Number of junior researchers* (3)	*Number of mid-career researchers* (4)	*Number of senior researchers* (5)	*Number of retiring researchers* (6)	*Share of women* (7)
All	6.294***	1.755***	0.246	0.710**	3.063***	0.899***	0.010
Std. error	(1.405)	(0.509)	(0.480)	(0.349)	(0.410)	(0.240)	(0.009)
Pre-treatment mean	22.290	4.746	9.095	4.507	4.829	2.315	0.230
N (units)	1,948	1,948	1,948	1,948	1,948	1,948	1,948
Controls							
Domain	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.E.4 : Effect of the Reform on Number of Researchers - Never Treated Only

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on the number of researchers in treated institutions per domain and per category: all (1), researchers who first published abroad (2), in the past 5 years (3), between 5 to 15 years prior to each year (4), and above 15 years (5), the number of researchers whose last publication was recorded this year (6) and the share of women (7). Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated units as controls. Each researcher with n>1 affiliations is counted as 1/n researcher at each affiliation. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Publications (1)	*Average publications* (2)	*Citations* (3)	*Average citations* (4)	*Top 5% journal publications* (5)	*Avg. Publications in top 5% cited journals* (6)	*Avg #new phrase combinations* (7)	*Avg. New phrases* (8)
All	13.369***	-0.155	35.641	7.081	1.710	0.022	-154.486*	7.218
Std. error	(3.460)	(0.102)	(112.069)	(5.487)	(1.379)	(0.060)	(89.534)	(7.566)
Pre-treatment mean	38.432	2.064	1490.831	82.925	17.464	0.939	2066.979	117.670
N (units)	1,948	1,948	1,948	1,948	1,948	1,948	1,948	1,948
Controls								
Domain	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CNRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.E.5 : Effect of the Reform on Productivity at the Institution Level - Never Treated Only

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy for treated institutions on total publications (1) and total citations (2) by researchers who became affiliated with the lab in the past 5 years, and total publications (3) and total citations (4) by researchers affiliated with the lab who first published abroad. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Publications (1)	*Top 5% journal publications* (2)	*Citations* (3)	*Total #new phrase combinations* (4)	*Avg #new phrase combinations* (5)
All	-1.068***	-0.346***	-28.544***	6.563	0.628
Std. error	(0.068)	(0.052)	(5.974)	(6.164)	(0.945)
Pre-treatment mean	3.473	2.415	202.140	220.596	44.296
N (units)	66,771	66,771	66,771	66,771	66,771
Controls					
Entry year (5y. bins)	Yes	Yes	Yes	Yes	Yes
Quartile of Pub. (selection)	Yes	Yes	Yes	Yes	Yes
Quartile of Cit. (selection)	Yes	Yes	Yes	Yes	Yes

Table 2.E.6 : Effect of the Reform on Productivity at the Individual Level - Never Treated Only

Notes: This table presents the Average Treatment Effect on the Treated of administrative autonomy on total publications (1), total citations (2), total publications in the top 5% most cited journals (3), total (4) and average (5) number of new phrase combinations weighted by their reuse on researchers whose affiliation in 2007 was subsequently treated. Results are obtained by estimation of Callaway and Sant'Anna (2021), using the never-treated units as controls. Variables are winsorised at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Annexe 3
Annexes du chapitre 3
The Life Cycle of Ideas

3.A — Topic Evolution Through Years

This section provides the list of topic trajectories for topics in economics in Figure 3.A.1 and chemistry in Figure 3.A.2. For those topics that follow a bell-shaped trajectory, it also provides the average trajectory around the peak (Figure 3.A.3) and after emergence (Figure 3.A.4) split by decade of peak.

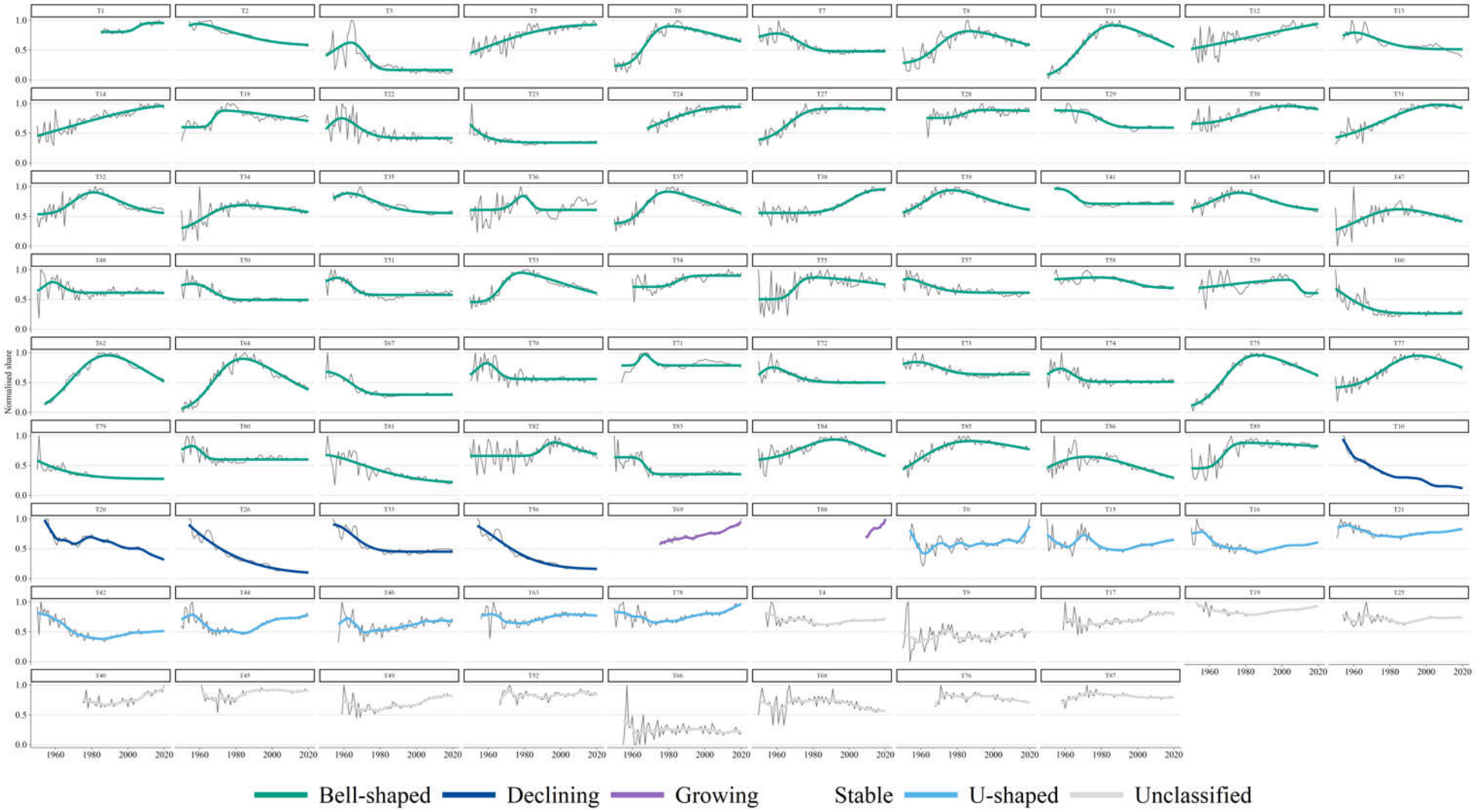


Figure 3.A.1 : Share of Publications and Trend in Economics, All Topics, Normalised Shares

The graph above represents the variation in the share of publications allocated to all 90 topics that we keep in our main analysis. The colour of the curve indicates the classification of the topic trajectory. The dotted vertical orange line represents the year of emergence of the topic. This version displays normalised shares.

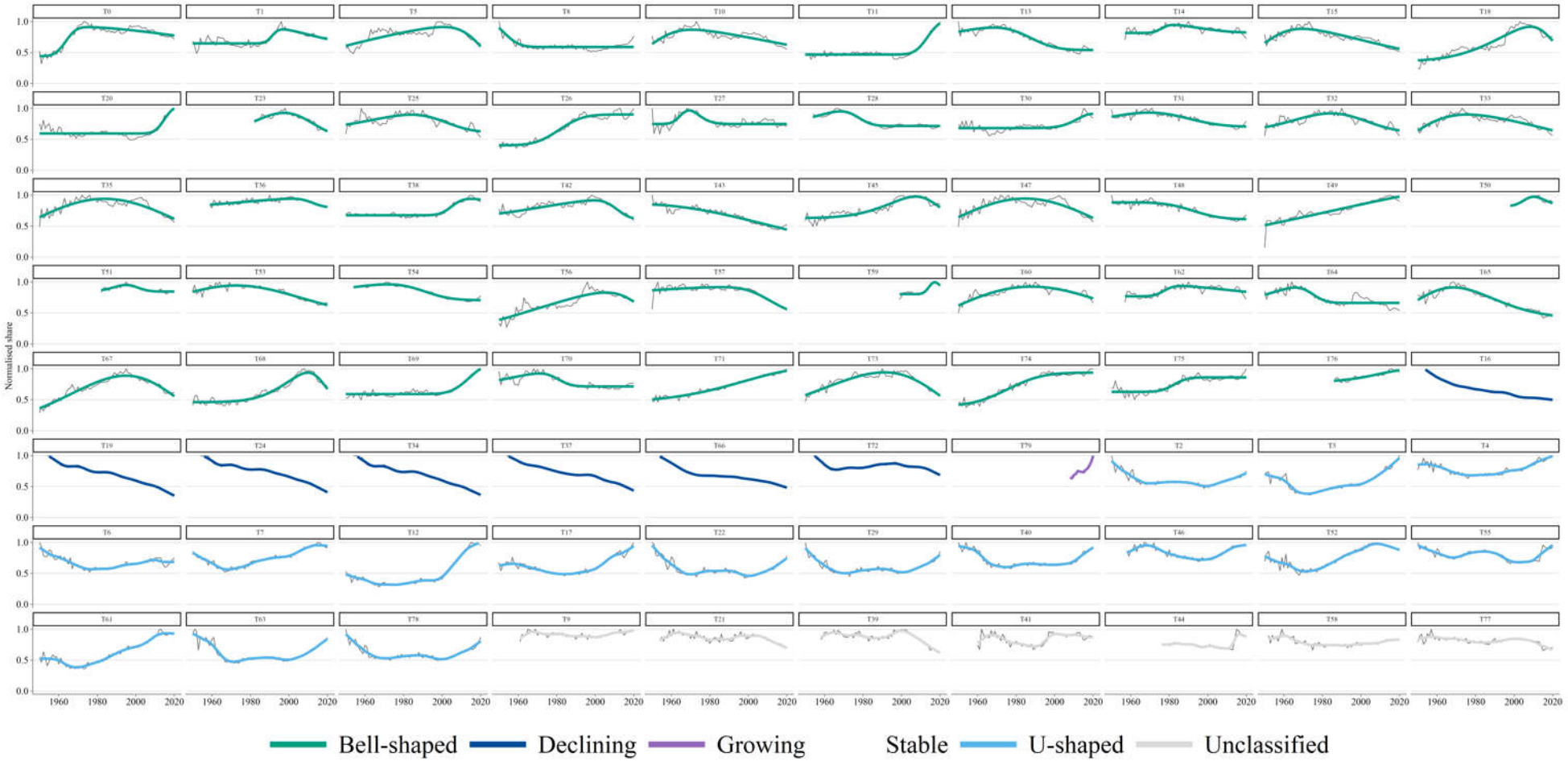
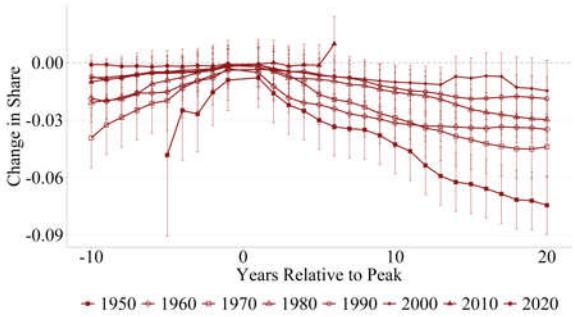


Figure 3.A.2 : Share of Publications and Trend in Chemistry, All Topics, Normalised Shares

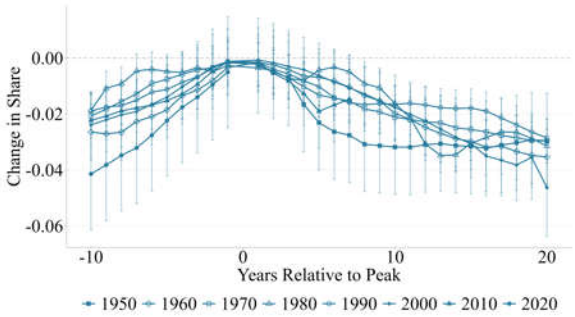
The graph above represents the variation in the share of publications allocated to all 80 topics that we keep in our main analysis. The colour of the curve indicates the classification of the topic trajectory. The dotted vertical orange line represents the year of emergence of the topic. This version displays normalised shares.

Figure 3.A.3 : Publication Share Trajectory Relative to Maximum, by Peak Decade

(a) Economics



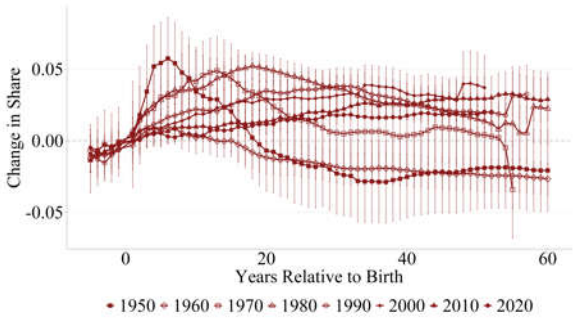
(b) Chemistry



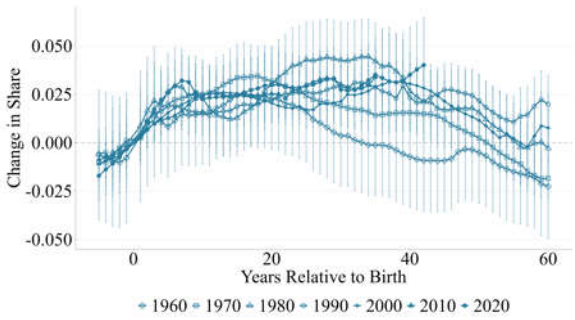
Notes: Each curve reports the estimated coefficients from a separate event-study regression of the smoothed top-10 publication share on relative-year indicators, estimated for each peak-decade cohort independently. Panel (a) plots the coefficients for the economics sample, while panel (b) plots the coefficients obtained with the chemistry sample. 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 10 years prior to the peak to 20 years after.

Figure 3.A.4 : Publication Share Trajectory Relative to Emergence, by Peak Decade

(a) Economics



(b) Chemistry



Notes: 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 emergence-decade cohort independently. Panel (a) plots the coefficients for the economics sample, while panel (b) plots the coefficients obtained with the chemistry sample. 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 emergence 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 emergence-year level. Topics are assigned to a decade cohort according to the calendar decade in which their publication emergence 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 5 years before and 60 years after the emergence date.

3.B — Topic Weights and Classification

1. Defining a Classification for Topics

As an example, for economics, 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 3.B.1 reports the distribution of topic weights for economics 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.

Table 3.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.

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 organised around substantive economic questions, while methodological topics often appear as secondary components of the paper's topic profile.

2. Topic Labels

Topics were labelled using ChatGPT. Prompts will be made available in the replication package. The following tables present all topic labels with the corresponding top 5 keywords.

Table 3.B.2 : Topic Labels and Keywords: economics

ID	Topic label	Top 5 keywords
0	Insurance Markets: Energy Risk	insur, strategi, market, factor, scale
1	Experimental Contracting	prefer, incent, uncertainti, experi, subject
2	International Monetary Economics	insur, system, time, structur, theori
3	Organizational Economics: Firm Governance	structur, data, product, theori, signific
4	Public Utility Economics	public, good, privat, water, fund
5	Open Economy Macroeconomics	monetari, fiscal, macroeconom, open, action
6	Asymptotic Econometrics	estim, paramet, distribut, asymptot, error
7	Demographic Economics: Life Cycle	time, popul, ag, occup, children
8	Applied Microeconometrics: Survey Regression	data, signific, distribut, regress, measur
9	Industrial Organization: Market Power	market, power, likelihood, time, substitut
10	Labor Economics: Employment Programs	job, signific, limit, benefit, data
11	Dynamic Econometrics	model, dynam, estim, empir, simul
12	Environmental Development Economics	forest, poverti, conserv, poor, evalu
13	Business Cycle Theory	theori, structur, empir, fluctuat, product
14	Behavioral Contract Theory	experi, prefer, contract, subject, incent
15	Economics of Education	educ, school, student, famili, univers
16	Social Economics	social, commun, inequ, union, societi
17	Experimental Economics	uncertainti, experi, behavior, contract, individu
18	Household Public Finance	incom, household, tax, food, taxe
19	Regulatory Political Economy: Federal Policy	polici, regul, nation, state, domest
20	Microeconomic Theory: Demand	demand, theoret, theori, framework, properti
21	Regulatory Political Economy: Intervention	polici, regul, nation, state, feder
22	Financial History: Policy Systems	time, theori, pattern, measur, market
23	Development Experiments	group, network, member, ineffici, aid
24	Structural Estimation	predict, assumpt, empir, structur, estim
25	International Finance	capit, invest, asset, human, flow
26	Corporate Governance	strategi, structur, signific, measur, degre
27	Environmental Economics	cost, emiss, subsidi, carbon, transact
28	Development Economics	sector, reform, mobil, countri, state
29	Project Finance: Infrastructure	project, feder, financ, power, regim
30	Trade Welfare Analysis	welfar, mobil, loss, search, entri
31	Electricity Markets	market, volatil, competit, inform, sale
32	Differentiated Products Demand	price, demand, produc, equilibrium, stock
33	Financial Systems	limit, system, time, market, structur
34	Applied Econometrics: Factor Models	data, variabl, method, empir, index
35	Agricultural Economics	agricultur, land, farmer, product, farm
36	Urban Economics	spatial, energi, urban, citi, hous
37	Decision Theory	function, util, choic, optim, individu
38	International Political Economy	countri, global, regul, domest, south
39	Labor Supply Modeling	suppli, adjust, labour, elast, aggreg
40	Health Economics	health, care, insur, content, women
41	Regulatory Political Economy: Policy Systems	polici, regul, nation, system, state
42	Institutional Political Economy: Property Rights	state, institut, structur, resourc, strategi
43	Energy Demand	system, consumpt, demand, monei, electr
44	Corporate Sustainability	environment, organ, practic, sustain, adopt
45	Industrial Organization: Firm Competition	firm, competit, profit, differenti, busi
46	Firm Dynamics	factor, data, empir, variabl, structur
47	Applied Econometrics: Heterogeneity	structur, variabl, method, aggreg, empir
48	Trade Policy	market, polici, scale, protect, domest
49	Mechanism Design Experiments	experi, prefer, mechan, subject, incent
50	Innovation Economics	innov, industri, technolog, knowledg, enterpris
51	Institutional Economics	institut, resourc, polit, right, law
52	Panel Data Econometrics	data, variabl, signific, method, survei
53	Discrete Choice Theory	choic, function, optim, util, individu
54	Energy Macroeconomics	oil, factor, market, strategi, cycl
55	Behavioral Public Economics	individu, regim, margin, function, incent
56	Political Economy: Policy Implementation	strategi, measur, signific, effort, practic
57	Labor Economics	wage, labor, employ, worker, gender
58	Applied Microeconometrics: Panel Regression	data, variabl, method, signific, factor
59	Monetary Economics: Transmission	time, rate, system, model, benefit
60	Business Cycles: Corporate Finance	busi, job, framework, strategi, empir
61	Panel Econometrics	hypothesi, causal, data, regress, panel
62	Structural Econometrics	model, dynam, applic, mechan, variabl
63	Oil and Business Cycles	empir, oil, theori, busi, structur
64	Structural Demand Estimation	model, function, estim, data, dynam
65	Mechanism Design	program, inform, qualiti, particip, formal
66	Financial Markets: Interest Rates	rate, signific, limit, theori, market
67	Institutional Political Economy: State Capacity	institut, resourc, polit, state, capac
68	Labor Market Search	job, model, market, theoret, framework
69	Asset Pricing	risk, return, stock, investor, portfolio
70	Natural Resources: Institutional Reform	natur, ga, capac, field, poor
71	Economic Growth	growth, output, converg, infrastructur, countri
72	Demographic Economics: Age Structure	time, popul, statu, ag, survei
73	Household Demography	time, popul, ag, children, peopl
74	Financial Regulation: Insurance	polici, insur, nation, structur, domest
75	Dynamic Stochastic Models	model, dynam, stochast, linear, equat
76	Public Goods Provision	effici, expenditur, alloc, spend, provid
77	Game Theory	equilibrium, agent, game, ration, player
78	Local Public Finance	govern, local, debt, climat, budget
79	Spatial Service Access	servic, locat, access, payment, provid
80	International Trade	export, trade, foreign, commod, agreem
81	Macroeconomics: Distribution and Demand	system, benefit, signific, rate, integr
82	Electricity Reform	reform, electr, monei, system, substitut
83	Theory of the Firm	theori, strategi, factor, size, busi
84	Time Series Econometrics	test, seri, forecast, trend, cointegr
85	Monetary Macroeconomics	rate, exchang, shock, inflat, real
86	Dynamic Theory: Regime Models	time, model, theoret, theori, altern
87	Production Economics	product, consum, technologi, input, differenti
88	Banking Crises	financi, bank, credit, crisi, fund
89	Applied Econometrics: Causal Inference	data, method, signific, distribut, variabl

Table 3.B.3 : Topic Labels and Keywords: chemistry

ID	Topic label	Top 5 keywords
0	DNA conformational modeling	activ, site, receptor, catalyt, speci
1	Polymer synthesis	surfac, spectroscopi, organ, level, devic
2	Environmental and biological chemistry: Metal accumulation	densiti, theori, frequenc, rotat, vibrat
3	Porous sorbent materials	phase, liquid, column, carbondioxyd, pressur
4	Thermal polymer degradation	drug, cancer, network, stress, deliveri
5	Ion solvation	materi, capac, cycl, color, cathod
6	Phytochemistry and environmental toxicology	carbon, content, air, fuel, pollut
7	Surface modification: Amorphous coatings	cation, anion, ionic, alkyl, neutral
8	Environmental and biological chemistry: Chemical exposure profiling	spectra, band, raman, mode, fragment
9	Optical spectroscopy	temperatur, crystallin, min, residu, pressur
10	Intermolecular interactions: Crystal packing	temperatur, separ, solut, pressur, ethanol
11	Cancer chemical biology	temperatur, synthes, pressur, diffract, crystallin
12	Electrochemical sensors	magnet, interact, nmr, arrai, spin
13	Molecular excited states	solut, level, synthes, aggreg, aqueou
14	Organic synthesis: Functional group chemistry	organ, plant, industri, area, technologi
15	Intermolecular interactions: Interaction energetics	cell, protein, peptid, tissu, cellular
16	Chemical thermodynamics	solut, aqueou, separ, remov, organ
17	Dielectric materials	reaction, complex, atom, theoret, coordin
18	Coordination chemistry: Metal complexes	synthesi, surfac, activ, organ, cyclic
19	Electrolyte solutions: Diffusion	detect, oil, analyt, level, standard
20	Energy materials chemistry	activ, speci, volum, diseas, diffus
21	Chemical crystallography	rate, kinet, optic, densiti, diffus
22	Environmental and biological chemistry: Contaminant detection	surfac, atom, length, spectroscopi, comput
23	Phase behavior	ion, sensor, angstrom, coordin, neutral
24	Electrolyte solutions: Dilution thermodynamics	reaction, rate, oxygen, oxid, air
25	Organic and coordination chemistry	atom, interact, site, bond, energi
26	Quantitative chemical analysis	oxid, temperatur, thin, electrochem, materi
27	Magnetic and vibrational spectroscopy	solut, salt, aqueou, sodium, chlorid
28	Density functional theory	atom, interact, surfac, layer, electron
29	Environmental and biological chemistry: Drug analysis	materi, synthesi, synthes, area, densiti
30	Surface modification: Surface deposition	surfac, adsorpt, adsorb, monolay, copper
31	Condensed matter physics: Mechanical properties	electron, solut, interact, spectroscopi, oxid
32	Organic synthesis: Carboxylic acid chemistry	solvent, methyl, ester, solubl, pore
33	Coordination chemistry: Coordination geometry	materi, synthesi, area, cycl, framework
34	Electrolyte solutions: Aqueous solubility	field, equat, electr, flow, devic
35	Organic reaction mechanisms	compound, assai, isol, inhibitor, biolog
36	Metal coordination compounds	metal, emiss, hybrid, dye, materi
37	Glass and polymer thermophysics	electron, charg, graphen, dimer, orbit
38	Electron microscopy and diffraction	interact, nmr, magnet, spectroscopi, electron
39	Heterocyclic synthesis	solut, separ, level, detect, chromatographi
40	Battery chemistry	crystal, chain, copolym, block, crystallin
41	Charge-transfer photophysics	surfac, electrochem, pattern, synthes, electron
42	Nuclear magnetic resonance	complex, ligand, nmr, neutral, atom
43	Theoretical quantum chemistry	light, absorpt, optic, spectral, level
44	Enzyme catalysis	speci, synthes, inhibit, biolog, amino
45	Metal speciation and redox	particl, nanoparticl, surfact, gold, silica
46	Photovoltaic materials	diffus, speci, optic, absorpt, volum
47	Uninterpretable Cluster	solut, phenol, kinet, separ, devic
48	Condensed matter physics: Thermal expansion	energi, cluster, proton, bond, theoret
49	Infrared and photoelectron spectroscopy	solut, residu, comput, surfac, ethyl
50	Fluorescence spectroscopy	synthesi, organ, materi, area, food
51	Liquid chromatography	interact, electron, coordin, nmr, magnet
52	Thin film materials	synthesi, isol, biolog, synthes, comput
53	Materials degradation kinetics	interact, magnet, domain, gel, spectroscopi
54	Condensed matter physics: Structural domains	synthesi, organ, area, light, industri
55	Environmental and biological chemistry: Biological chemical profiling	state, energi, absorpt, spectra, fluoresc
56	Polymer chemistry	simul, dynam, fluid, quantum, theori
57	Organic stereochemistry	temperatur, celsiu, thermal, heat, silver
58	Oxidation catalysis	materi, area, resist, synthesi, light
59	Heterogeneous catalysis	film, layer, electrod, deposit, electrolyt
60	Organic synthesis: Substitution reactions	solut, min, kinet, phosphat, organ
61	Carbon material interfaces	activ, catalyt, site, enzym, vitro
62	Organic synthesis: Cyclic compounds	atom, electron, spectroscopi, pair, interact
63	Environmental materials chemistry	densiti, temperatur, speci, coeffici, liquid
64	Hydrogen bonding	microscopi, electron, nanotub, morphologi, fiber
65	Absorption spectroscopy	synthes, spectroscopi, infrar, diffract, magnet
66	Organic functional group chemistry	reaction, alcohol, reactiv, nucleophil, amin
67	Chromatography and mass spectrometry	reaction, rate, kinet, reactiv, atmospher
68	Colloid chemistry	fluoresc, state, dna, excit, absorpt
69	Chemical biology	surfac, spectroscopi, level, absorpt, optic
70	Ab initio quantum chemistry	acid, acet, sulfur, content, min
71	Protein and peptide chemistry	temperatur, solut, pressur, separ, coeffici
72	Liquid mixture thermodynamics	oxid, catalyst, oxygen, iron, reactiv
73	Nucleophilic substitution	polym, polymer, weight, monom, radic
74	Molecular dynamics simulation	mass, spectrometri, plasma, ioniz, inject
75	Fluid transport	pressur, temperatur, degrad, separ, rate
76	Adsorption and membrane separation	speci, activ, synthesi, surfac, host
77	Metal ion chemistry	bond, hydrogen, atom, plane, coordin
78	Environmental and biological chemistry: Environmental toxicology	membran, polar, radic, lipid, voltag
79	Medicinal chemistry	interact, conform, state, electrostat, theoret

1. Defining a Classification for Trajectories

Each of the topics is assigned to one of six potential categories: *bell-shaped*, *declining*, *growing*, *stable*, *U-shaped*, or *unclassified*. Classification is performed using an algorithm that fits a Local Estimated Scatterplot Smoothing (LOESS)-smoothed curve to each topic's annual top-10 share 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. Overall, no topic in either discipline is classified as stable by our method, which is why we drop it in the main text.

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 in the normalised series) ensures that only well-defined peaks are accepted.

Round 2 — Smoothed series, trend detection enabled for topics that do not follow a bell shape. 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 weights 0.1, 0.1, 0.2, 0.3 and 0.3 from the oldest year to the current one). 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, allowing the 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-emergence window. Remaining unclassified topics undergo a final pass on the raw series, now restricted to years at or after each topic's estimated year of emergence. This filter removes the pre-emergence 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. A topic must not feature a dip below 0.15 of its minimum value in order to be considered as growing. It also activates detection of U-shapes. The prominence threshold 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 not unimodal.

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 normalised 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.

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

Group	Parameter	Round 1	Round 2	Round 3
Input series	Variable	raw	**smoothed**	**raw, from emergence**
	Pre-smoothing	none	**5-yr BWA**	**none**
Smoothing and bell test	`bw` (LOESS span)	0.20	0.20	0.20
	`tol` (bell tolerance)	0.30	0.30	0.30
Peak and unimodality	`edge` (yr)	0	0	0
	`thrmon` (monotone share)	0.30	0.30	0.30
	`thprom` (peak height)	0.15	0.15	**0.08**
Declining	`show_trend`	FALSE	**TRUE**	TRUE
	`trend_edge` (yr)	—	5	5
	`trend_thr_down`	—	0.60	0.60
Growing	`show_growth`	FALSE	FALSE	**TRUE**
	`grow_edge` (yr)	—	—	5
	`grow_thr_up`	—	—	0.60
	`grow_max_dip`	—	—	0.15
U-shaped	`show_u`	FALSE	FALSE	**TRUE**
	`u_edge` (yr)	—	—	5
	`u_thrmon`	—	—	0.65
	`u_prom`	—	—	0.10
Stable	`show_stable`	TRUE	TRUE	TRUE
	`stable_range`	0.10	0.10	**0.05**
	`stable_slope` (per yr)	0.002	0.002	**0.50**

Notes. Each round classifies only the topics left unclassified by the previous round. Each topic's series is divided by its maximum, so all thresholds are on a 0–1 scale, and is smoothed by LOESS (degree 1). Categories are assigned in this order of priority: stable, declining, bell-shaped, U-shaped, growing; remaining topics are unclassified. **Input series**: share of papers in which the topic has one of its ten highest weights. Round 2 pre-smooths it with a backward-weighted moving average (BWA) with weights $(0.1, 0.1, 0.2, 0.3, 0.3)$ on years $t-4$ to t. Round 3 uses the series from the topic's emergence year onwards, for topics that emerge by 2010. **bw**: share of observations in each local LOESS window; larger values give smoother curves. **tol**: a topic is bell-shaped if a skew-normal curve fits its smoothed series nearly as well as a cubic polynomial, $\mathrm{RSS}_{\mathrm{bell}} \leq (1 + \texttt{tol})\,\mathrm{RSS}_{\mathrm{cubic}}$. Only topics that pass the unimodality test are fitted, and fits require at least 20 years. **edge**: the peak must lie at least `edge` years from both ends of the series; with `edge` = 0, a topic may peak in the first or last year. **thrmon**: at least this share of year-on-year changes must be increases before the peak and decreases after it. **thprom**: the peak must exceed by at least `thprom` the median level of the first and last 20% of years. **Declining**: the peak falls within the first `trend_edge` years, and at least a share `trend_thr_down` of year-on-year changes after it are decreases. **Growing**: the peak falls within the last `grow_edge` years, at least a share `grow_thr_up` of year-on-year changes before it are increases, and before the peak the series never falls below its previous maximum by more than `grow_max_dip` times its range. **U-shaped**: the minimum lies at least `u_edge` years from both ends, at least a share `u_thrmon` of year-on-year changes are decreases before it and increases after it, and the median levels of the first and last 20% of years both exceed it by at least `u_prom`. **Stable**: the range of the smoothed series is at most `stable_range` and the absolute slope of a linear trend is at most `stable_slope` per year; in Round 3 the slope condition does not bind, so only the range applies. —: not used because the feature is switched off. Values in bold differ from the previous round.

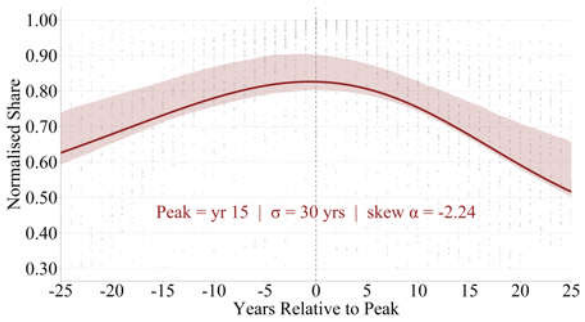
The model is then fitted jointly on all (i, t) pairs, in which t denotes years relative to the publication peak, rather than on pre-aggregated annual means. Formally, we fit the following skew-normal specification:

$$y_{it} = \frac{b_0}{\sigma} + e^{A} \cdot 2\,\phi\left(\frac{t-\mu}{\sigma}\right)\Phi\left(\alpha \cdot \frac{t-\mu}{\sigma}\right) + \varepsilon_{it} \tag{3.C.1}$$

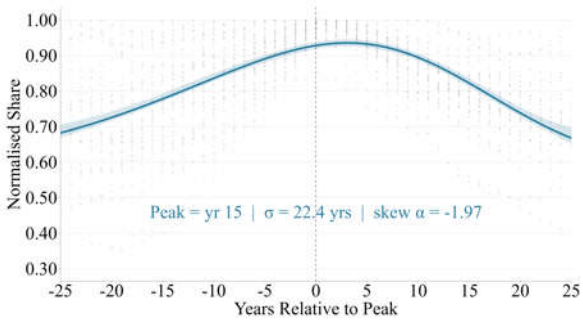
where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and cumulative distribution function, respectively. The parameter μ captures the location of the peak relative to year zero (the identified maximum), σ determines the width of the bell in years, α controls the degree and direction 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 3.C.1 : Aggregate Trend, All Topics

(a) Economics



(b) Chemistry



Notes: Each grey dot represents one topic-year observation, where the publication share series of each topic has been normalised by its own within-topic maximum so that the peak equals one. The red and blue curves, for economics and chemistry respectively, show 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.

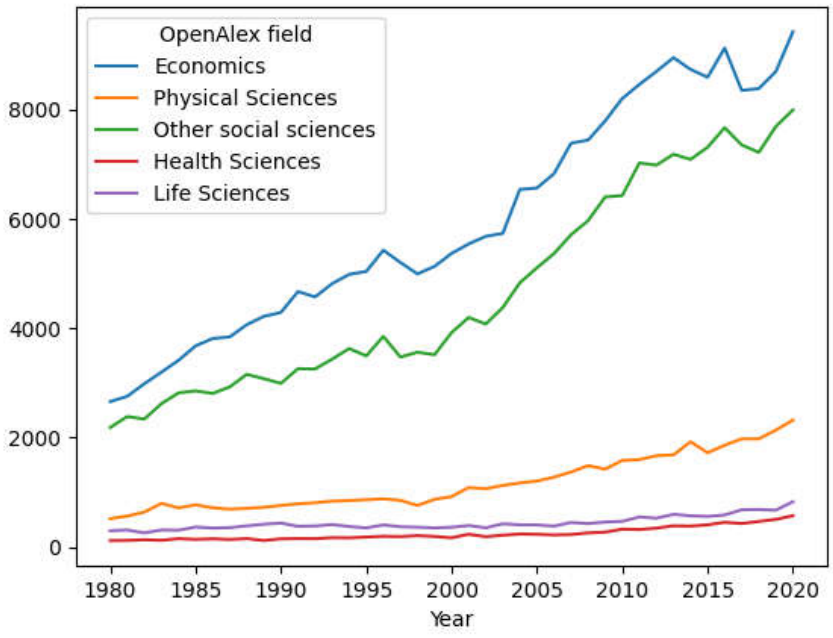
3.D — Data Collection and Treatment

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 in economics) 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 optimisation. Due to this, some topics are too small, such as Economic Burden of Cancer Treatment, while 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 or chemistry. For economics, Figure 3.D.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 for the authors who published in our sample is even more skewed away from economics, as shown in Figure 3.D.2. Papers tagged as economics in OpenAlex are a smaller portion of the publications of these authors.

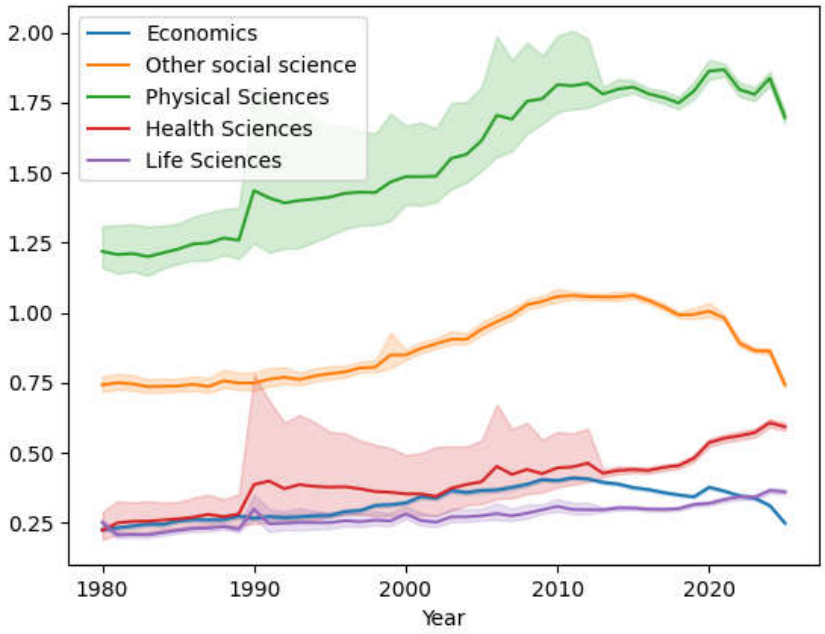
Figure 3.D.1 : Number of Total Publications in Sample, Per Domain



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.

We also provide in Figure 3.D.4 the same graph as pane (b) of Figure 3.D.3 and Figure 3.D.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.

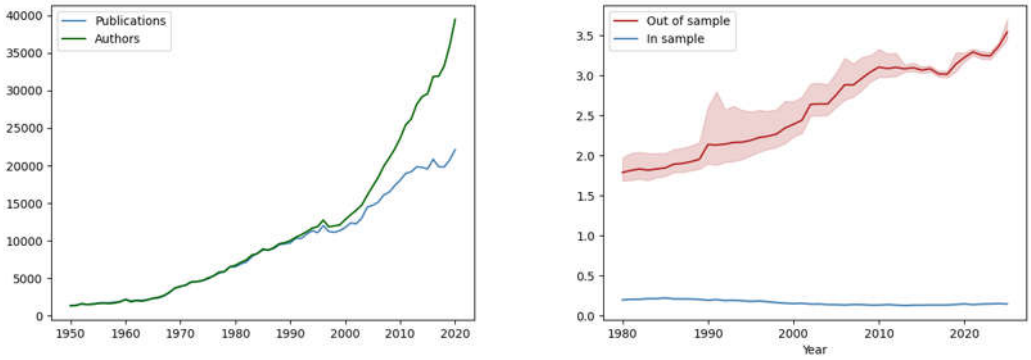
Figure 3.D.2 : Average Number of Publications per Year per Researcher, by Domain



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 3.D.3 : Total number of entries in sample and in entire database

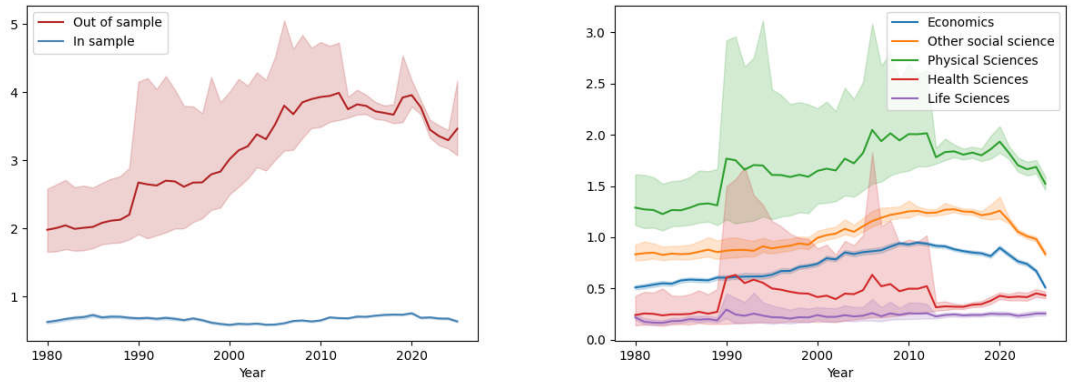
(a) In sample (b) In the entire database



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 3.D.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.

3.E — Details on Topic Identification

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 modelling we use is Latent Dirichlet Allocation (Blei, Ng, and Jordan, 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.[113]

The logic behind LDA is not limited to the analysis of semantic content: it is a likelihood maximisation 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 (i.e., pronouns, "is", "the", "of"),
- Rare words (those that appear too infrequently),
- Ubiquitous words (those that appear in nearly all documents).

2. Embedded Topic Model

The Embedded Topic Model (Dieng, Ruiz, and Blei, 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 logistic-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.

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.[114]

3. Algorithmic Performance Metrics

a) Log-likelihood

The most obvious metric for the evaluation of a model is its target; here, it is 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

[113] 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 tractable lower bound on it.

[114] *Idem.*

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.

b) **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 exponential of minus the average per-word log-likelihood on held-out documents. If the data generating process of the model also applies to the out-of-sample documents, then the model is generalisable, 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, Ng, and Jordan (2003) define perplexity as:

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

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.

3.F — Optimisation of the Embedded Topic Model

1. Optimisation

In practice, the embedded topic model (Dieng, Ruiz, and Blei, 2020) to be estimated on a dataset of documents needs 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 fewer than a given share of documents and above a given share of documents. As mentioned above, we iterate in order to find the specification that fits our data best. The combinations of the following parameters are considered: 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%.

Because estimation time grows with the number of topics and the size of the vocabulary, we draw a random sample of 20% of the abstracts, stratified by year, prior to applying the algorithm and split it into a training set (80% of the subset) and test set (20% of the subset) which are the same for all specifications. All specifications are trained for 100 epochs.

Fit is then measured by document-completion perplexity on the held-out test split, as implemented in the ETM package v.1.2.1: topic proportions are inferred from the first half of each test abstract, and the per-word log-likelihood of the second half is averaged across abstracts and exponentiated. The kink is then determined by visual inspection of the perplexity curve. Once the model is determined, it is applied without re-estimation to the entire corpus for each discipline.

2. Robustness and Validation on Simulated Data

Placebo Bell Curves

The shares of topics within a paper must always sum up to one. This built-in property of the ETM could be driving the trajectories that we observe in the data: as some topics gain ground, others must fall to compensate. We confirm that this is not the case by running a first placebo exercise. We simulate 1,000 random datasets in a similar fashion to above, using economics as a benchmark for both disciplines: for each observed year, the datasets contains as many abstracts of length L as there are in the economics corpus. These abstracts are generated from a single, time-invariant word distribution, and their topics are assigned with the economics model. We then classify the trajectory of topics in these random datasets exactly as we do for observed data. Figure 3.F.1 displays the results of this exercise: in these 1,000 simulated datasets, on average, our procedure classifies 10% of topics as bell-shaped. In contrast, we assess that in observed data, topics given by the model follow a bell-shape in about two thirds of cases, for both disciplines.

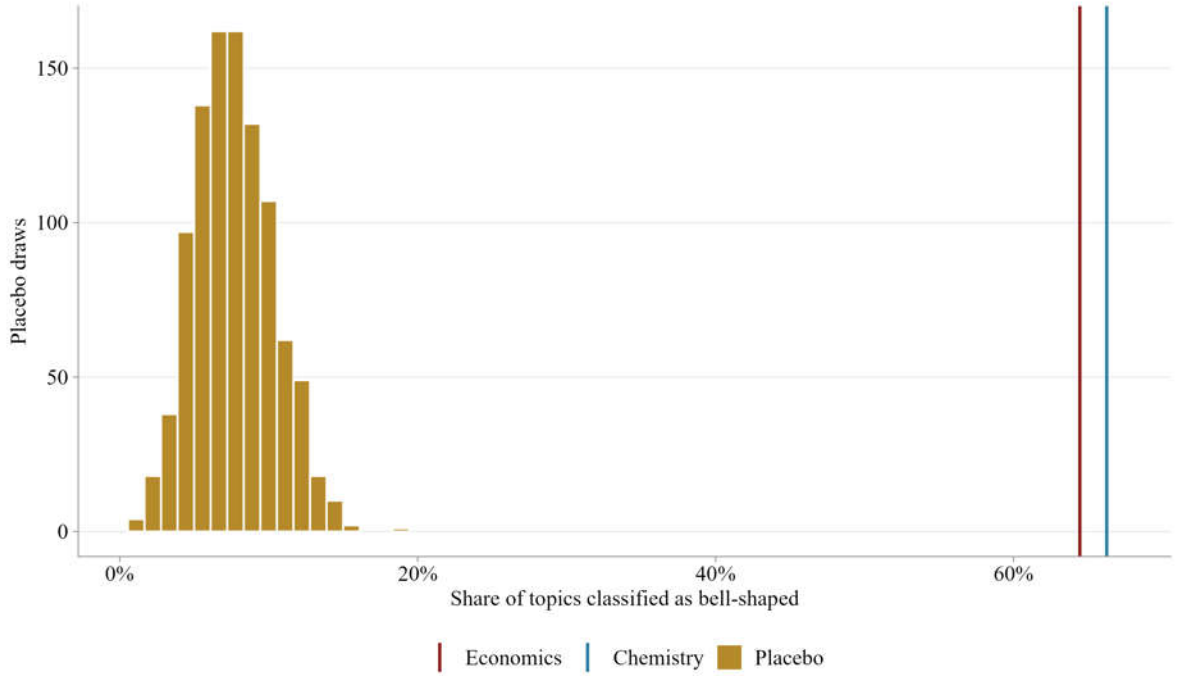


Figure 3.F.1 : Bell-shaped topics in observed and simulated data

Notes: Share of topics classified as bell-shaped in economics (red line) and chemistry (blue line), and its distribution across 1,000 simulated datasets (yellow histogram). Each simulated dataset contains, for every year and abstract length, as many abstracts as the economics corpus. Its abstracts are drawn from a single time-invariant word distribution, the average of the economics topic–word distributions, so that no topic rises or falls by construction. Topics are then assigned with the estimated economics model and classified as in the observed data.

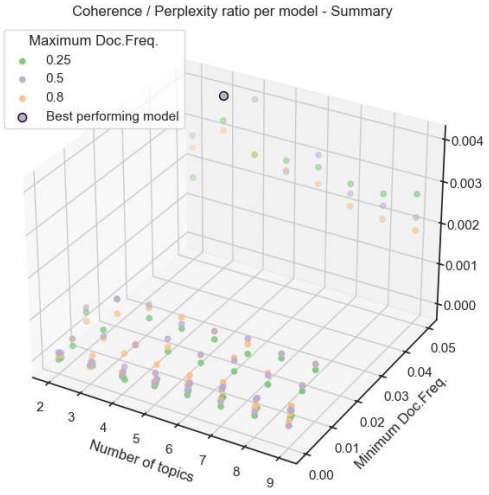


Figure 3.F.2 : Coherence to perplexity ratio for all specifications in the three-subjects simulated data

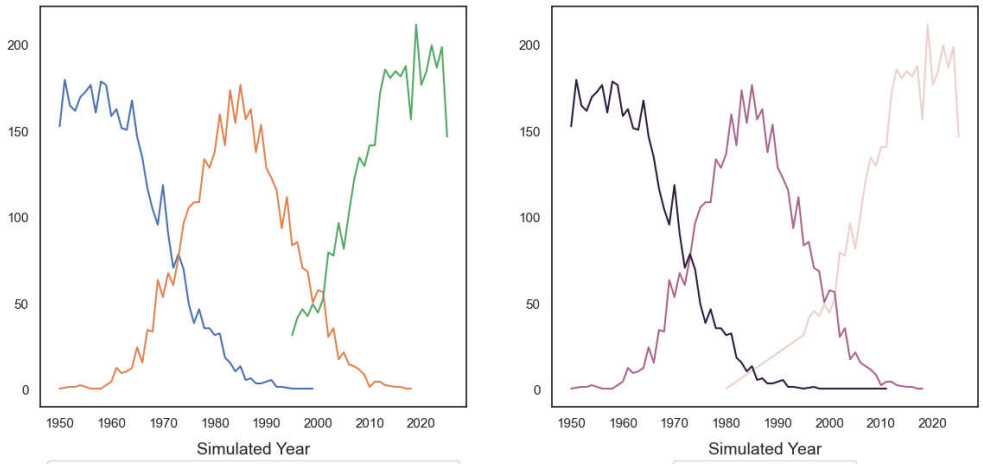
Placebo subjects

In order to assess the effectiveness of our optimisation process in choosing the right number of topics for a dataset, we perform the entire algorithm on 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 II.II.1, which leaves us with 11,846 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 Section II.II.2. Figure 3.F.2 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 almost perfectly replicated as shown in Figure 3.F.3.

Figure 3.F.3 : 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.

3.G — Alternative measures

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

Our alternative measure comes from Arts, Melluso, and Veugelers (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 3.G.1 shows that unlike 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 field generality[115] in papers in economics at the field level.

[115] We do not measure generality with respect to fields of research in this paper.

Figure 3.G.1 : Average Semantic Distance to all OpenAlex Papers

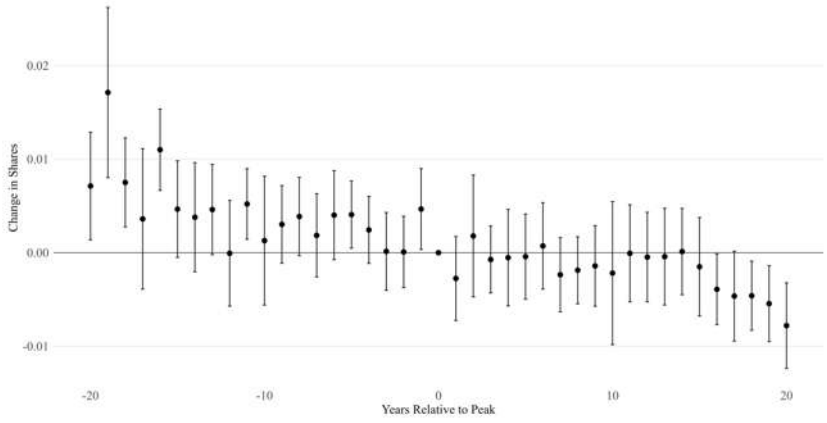


TABLE DES MATIÈRES

REMERCIEMENTS . . . . . **iii**

RÉSUMÉ ET MOTS CLÉS . . . . . **v**

ABSTRACT AND KEYWORDS . . . . . **vii**

SOMMAIRE. . . . . . **ix**

TABLE DES TABLEAUX . . . . . **xi**

TABLE DES ILLUSTRATIONS . . . . . **xiii**

GENERAL INTRODUCTION **1**

CHAPITRE 1 — HOW TO SILENCE RESEARCHERS? EVIDENCE FROM ILLIBERAL POLICIES IN HUNGARY **9**

I — Introduction 9
II — Political Context, Illiberalism and Threats to Academic Freedom 13
III — Academic Data Collection 15
1. Current Research Information Systems *15*
2. International Bibliometric Data *16*
3. Measures of Academic Performance and Mobility *17*
a) Quantity, Quality and Network 17
b) Research Direction 17
c) Researcher Mobility 18
IV — Empirical Strategy 18
1. Sample Selection. *18*
a) Hungarian Scholars and Political Targets 18
b) Hungarian and Central-Eastern European Researchers 21
2. Comparing Academic Performance Across Researchers *21*
V — How Do Political Pressures Affect Scientists' Productivity? 22
1. The Effect of Being Perceived as an Opponent on Academic Productivity *23*
a) Perceived Political Opponents 23
b) Targeted Topics 25
2. Mechanisms of Repression *26*
a) Is it Just Funding?. 26
b) Network Shrinking Effects 28
c) Loss of Influence 29
VI — Academic Strategies under Political Pressure. 30
1. Theoretical Framework *31*
2. Exiting Hungary *31*
3. Voicing Dissent *32*
4. Between Loyalty and Silence: Self-censorship *36*
VII — What Is the Global Effect of Illiberal Policies? 39
1. Comparing Individual Research Outcomes between Countries *39*
2. Research Productivity and Research Quality *39*
3. Decoupling of National and International Collaboration Networks *40*
4. Exit Through International Mobility and Leaving Academia *41*
5. Aggregate Losses from Political Pressure. *42*
VIII — Conclusion 46

CHAPITRE 2 — CHARACTERISING THE GAINS TO AUTONOMY IN THE PUBLIC SECTOR: EVIDENCE FROM FRENCH ACADEMIA 47

I — Introduction 47
II — Context and Content of the Reform 51
III — Data and Methodology 53
1. Data Collection 53
2. Selection of Samples 53
3. Empirical Strategy 55
IV — How Did the Reform Affect Institutions? 56
1. Autonomy Increased the Stock of Researchers 56
2. Autonomy Increased Total but not Average Productivity 59
3. Heterogeneity by IDEX Status and Field 60
V — How Did the Reform Affect Researcher Productivity? 62
1. Did the Reform Change Researcher Productivity? 62
2. Who Reaps the Gains of Autonomy? 64
VI — Discussion 65
1. Robustness. 66
2. External Validity and Aggregate Implications 67
VII — Conclusion 68

CHAPITRE 3 — THE LIFE CYCLE OF IDEAS 69

I — Introduction 69
II — Data and Method. 71
1. Data Acquisition 71
2. Allocating Papers to Topics 73
III — Implementation of the Algorithm 76
IV — Stylized Facts on the Lifecycle of Ideas in Economics and Chemistry 77
1. The Topic Life Cycle: Evidence of a Rise-and-Fall Pattern 77
2. Global Growth in Scientific Ideas 79
3. Reset of the Innovation Clock 82
V — Conclusion 84

BIBLIOGRAPHIE 85

Ouvrages imprimés 85
Chapitre dans un ouvrage imprimé 86
Rapports imprimés 86
Travaux universitaires 88
Articles de périodiques imprimés 89
Communication dans un congrès 96
Autres sources 97

ANNEXE 1 — ANNEXES DU CHAPITRE 1 : HOW TO SILENCE RESEARCHERS? EVIDENCE FROM ILLIBERAL POLICIES IN HUNGARY 98

1.A — Political Background and the State of Academia in Hungary 98
1. Descriptive Evidence of the Impact of Illiberalism on Hungarian Academia 98
2. The Central European University 101
1.B — Data Collection and Harmonization Process 104
1. Collection of the Data Sets 104
2. Harmonization of Fields. 105
3. Merging National and International Data 106
a) Matching Publications 106
b) Matching Authors 106
c) Eliminating Incorrect Matches 107
4. Comparing Academic Trajectories across Central European Countries 108
1.C — Academics' Political Alignment 109
1. Perceived Political Opponents 110
2. Perceived Political Allies 111
3. Networks of Opponents, Networks of Allies. 113
a) Network Description 113
b) Heterogeneity Results 115
4. Targeted Research Topics 116
a) Topic Definition 116
b) Gender 116

c) Rule of Law . . . 117
d) Sensitive Topics . . . 117
e) Climate change . . . 118
f) Descriptive Statistics on New Researchers . . . 120
1.D — Theoretical Framework: Exit, Voice, Loyalty or Self-censorship . . . 121
1. Setup . . . *121*
2. Solving the Model . . . *121*
a) The Frontiers . . . 122
b) Partition of the (m, θ) Plane . . . 122
c) Voice Confined to Intermediate Pressure . . . 123
d) Loyalty first, Voice, then Silence . . . 124
3. Predictions . . . *125*
4. Population Shares . . . *126*
1.E — Self-censorship and Research Direction . . . 129
1. Measuring Research Similarity . . . *129*
2. Similarity to one's own past research . . . *130*
3. Similarity to allies and opponents . . . *130*
4. Descriptive statistics . . . *130*
5. Publication Quality and Journals . . . *131*
a) Journal Rank . . . 131
b) Journal-level Persistence . . . 132
1.F — Political Alignment and Impact on Productivity: Complementary Results . . . 133
1. Political Opponent, Political Ally . . . *133*
2. Impact on the Members of the Networks of Politically Aligned Researchers . . . *135*
3. Targeted Topics: Climate . . . *135*
4. Targeted Topics for Political Opponents. . . . *137*
5. Channels of repression . . . *137*
a) Funding and the Effect of Illiberal Policies. . . . 137
b) International Network and Loss of Influence . . . 143
1.G — *Exit* and *Voice*: Complementary Results . . . 147
1. Voicing . . . *147*
2. Ally Network . . . *148*
3. Political Opponent Network . . . *150*
4. Additional Results on Research Similarity and Self-censorship . . . *151*
1.H — Global effect: Complementary Results . . . 154
1. Migration . . . *154*
1.I — Robustness Exercises . . . 156
1. Within Hungary: Opponents and Allies . . . *156*
2. Cross-Country: Hungary vs. Slovenia and Croatia . . . *161*

ANNEXE 2 — ANNEXES DU CHAPITRE 2 : CHARACTERISING THE GAINS TO AUTONOMY IN THE PUBLIC SECTOR: EVIDENCE FROM FRENCH ACADEMIA . . . 165

2.A — Appendix to Section II . . . 165
2.B — Appendix to Section III . . . 166
1. Sample selection . . . *166*
a) Institution-level Dataset . . . 166
b) Researcher-level Dataset . . . 166
2.C — Appendix to Section IV . . . 170
2.D — Appendix to Section V . . . 179
2.E — Appendix to Section VI . . . 180
1. Varying the Event Study Designs . . . *180*
a) Without Controls . . . 180
b) Using Only the Never-Treated as a Comparison . . . 181
c) Estimates by Groups from Core Regressions . . . 181
2. Varying the Year in which Treatment is Assigned . . . *182*

ANNEXE 3 — ANNEXES DU CHAPITRE 3 : THE LIFE CYCLE OF IDEAS . . . 186

3.A — Topic Evolution Through Years. . . . 186
3.B — Topic Weights and Classification. . . . 189
1. Defining a Classification for Topics . . . *189*
2. Topic Labels . . . *191*
3.C — Topic Popularity Trajectories . . . 194
1. Defining a Classification for Trajectories . . . *194*
2. Bell-shaped Trajectories . . . *194*
3.D — Data Collection and Treatment . . . 197

3.E — Details on Topic Identification 200
1. Latent Dirichlet Allocation *200*
2. Embedded Topic Model *200*
3. Algorithmic Performance Metrics. *200*
a) Log-likelihood 200
b) Perplexity 201
3.F — Optimisation of the Embedded Topic Model 201
1. Optimisation *201*
2. Robustness and Validation on Simulated Data *202*
3.G — Alternative measures 204

TABLE DES MATIÈRES 206