

Qui bénéficie le plus de la collaboration scientifique : la France ou les États-Unis ?

Bouzid Nélia, Braud-Noailly Angèle, Gariazzo Emilie, Lesauvage Leslie, Valdecasa Aurélien

I. Introduction

Dans un paysage scientifique mondialisé, la collaboration internationale est devenue un levier indispensable au rayonnement des chercheurs. Toutefois, cette architecture mondiale est structurée par des pôles de centralité, au premier rang desquels figurent les États-Unis. En raison de leur concentration de ressources financières et de leur rôle moteur dans l'innovation, les institutions américaines occupent une position hégémonique qui influence les trajectoires de recherche à l'échelle globale.

Dès lors, il convient d'interroger la réciprocité des bénéfices tirés de ces échanges transatlantiques : la collaboration universitaire entre la France et les États-Unis génère-t-elle un rendement scientifique asymétrique, et, dans quelle mesure le profit tiré de ce partenariat en termes de visibilité internationale est-il plus significatif pour les chercheurs français ou pour leurs homologues américains ?

II. Hypothèses

Notre hypothèse principale est que les chercheurs français bénéficient davantage de la collaboration franco-américaine que les chercheurs américains. Nous décidons de mesurer ces bénéfices par le nombre et la diversité des collaborations postérieures à cette première collaboration.

Hypothèse 1 : Les chercheurs français qui collaborent à l'international mais pas avec les USA ont moins de collaborations à l'international que les chercheurs français ayant collaboré avec des chercheurs américains.

Hypothèse 2 : Après une première collaboration universitaire entre la France et les États-Unis, les chercheurs français voient leur nombre de collaborations internationales augmenter de manière plus significative que celui des chercheurs états-uniens après une première collaboration avec la France.

Hypothèse 3 : Les chercheurs français diversifient davantage leurs collaborations internationales après leur première collaboration franco-états-unienne que les chercheurs américains.

Nous pensons effectivement que la collaboration avec les États-Unis est perçue comme prestigieuse pour les chercheurs français, en raison des moyens financiers dont bénéficient les universités américaines et de la place qu'occupent les États-Unis dans l'innovation et la production scientifique. Dans cette perspective, la collaboration avec des chercheurs américains constituerait un marqueur d'intégration aux réseaux académiques internationaux les plus centraux.

Selon ce raisonnement, les chercheurs français ayant collaboré avec des chercheurs américains bénéficieraient d'un accès privilégié à ces réseaux, ce qui se traduirait par un nombre plus élevé de

collaborations internationales que celui des chercheurs français collaborant à l'international sans jamais collaborer avec les États-Unis (Hypothèse 1).

Une première collaboration entre un chercheur français et un chercheur américain permettrait en outre de mettre en lumière le travail du chercheur français, d'abord au sein de l'espace académique national, puis à l'international, favorisant ainsi une augmentation plus marquée de son nombre de collaborations universitaires internationales. Cet effet serait plus limité pour les chercheurs américains, dont les trajectoires de collaboration sont déjà fortement internationalisées (Hypothèse 2).

Enfin, la forte mobilité des chercheurs du monde entier vers les États-Unis conduit à penser que les chercheurs américains sont, en moyenne, déjà habitués à collaborer avec des partenaires de nationalités diverses. À l'inverse, une première collaboration franco-américaine pourrait constituer pour les chercheurs français un point de bascule plus important, les incitant à diversifier davantage leurs collaborations internationales. Cette dynamique pourrait également s'expliquer par le fait que les chercheurs français chercheraient prioritairement à collaborer avec des universitaires américains avant d'élargir leurs collaborations à d'autres pays (Hypothèse 3).

III. Collecte et construction de la base de données

Les données utilisées dans cette étude ont été collectées à partir de la base bibliométrique ouverte OpenAlex, qui recense des publications scientifiques, leurs auteurs ou encore les institutions auxquelles elles sont rattachées. La période retenue s'étend de 1990 à 2025, ce qui permet d'analyser les trajectoires de collaboration sur le long terme et d'observer l'évolution des réseaux scientifiques avant et après les premières collaborations franco-américaines.

Pour réaliser notre étude, nous nous sommes concentrés sur la collecte de données de chercheurs français et américains. La première étape a donc été de récolter les métadonnées de chercheurs issus de quatre universités françaises présentant des niveaux différenciés de prestige et de visibilité internationale, tels qu'indiqués par le classement Shanghai ARWU 2025 : PSL (Paris Sciences & Lettres, 34e dans le monde et 2e en France), Sorbonne Université (43e dans le monde et 3e en France), l'Université de Montpellier (151-200e dans le monde et 6-8e en France) et l'Université Toulouse Capitole (401-500e dans le monde et 15-18e en France).

Le cas de PSL a requis un traitement particulier, dans la mesure où cette université n'a été créée qu'en 2012 et regroupe plusieurs établissements préexistants. Afin d'assurer la cohérence temporelle des trajectoires de publication observées sur l'ensemble de la période étudiée, nous avons fait le choix de définir un périmètre institutionnel stabilisé, fondé sur une sélection explicite des établissements composant PSL. Ont ainsi été inclus dans l'analyse le Collège de France, l'École normale supérieure, l'École nationale des chartes, l'École pratique des hautes études, l'ESPCI Paris, Chimie ParisTech et ses unités de recherche directement rattachées (Institut de recherche de Chimie Paris), l'Institut Curie, l'Observatoire de Paris, l'École nationale supérieure des Mines de Paris et Paris-Dauphine.

En revanche, plusieurs établissements affiliés à PSL ne purent être intégrés à l'analyse. Aucun auteur ne put être identifié pour le Conservatoire national supérieur d'art dramatique – PSL ni pour l'École nationale supérieure d'architecture Paris-Malaquais – PSL, dont l'entrée officielle au sein de PSL n'intervint qu'en 2025. De même, l'Unité de technologies chimiques et biologiques pour la santé (UTCBS), rattachée à Chimie ParisTech, fut exclue en raison de son affiliation institutionnelle à l'Université Paris Cité. L'Institut de recherche et développement sur l'énergie photovoltaïque ne pu être

retenu, faute d'identification fiable dans les bases exploitées. Enfin, pour l’Institut Curie, le périmètre fut volontairement limité à l’institution elle-même, sans distinction interne entre ses différentes structures, ce filtrage étant jugé suffisant au regard des objectifs de l’étude.

Pour chacune de ces institutions françaises, nous avons identifié un ensemble de chercheurs ayant collaboré avec des chercheurs américains avant l’année 2020. Ce seuil temporel a été retenu afin de disposer d’une période suffisante pour analyser les collaborations internationales postérieures à cette première collaboration franco-américaine. Aucun filtre n’a été appliqué aux institutions américaines partenaires, l’objectif étant de capter l’ensemble des collaborations franco-américaines sans restriction a priori.

La collecte des données a été réalisée à l'aide d'un script Python exploitant l'API d'OpenAlex. Chaque institution française a fait l'objet d'une collecte distincte. La procédure s'est déroulée en deux étapes. Dans un premier temps, le script a permis d'identifier l'ensemble des auteurs dont la dernière affiliation institutionnelle connue correspond à l'une des institutions françaises sélectionnées et de récupérer des métadonnées comme le nom, la dernière institution à laquelle ils sont rattachés ou encore leur identifiant OpenAlex. Dans un second temps, l'intégralité des publications associées à chacun de ces auteurs a été récupérée.

Pour chaque publication, le script extrait les principales métadonnées, telles que : l'identifiant de l'auteur à l'origine de la requête, son nom, l'identifiant unique du papier, la thématique et le champ de recherche, la liste des co-auteurs, l'année de publication, la dernière institution connue de l'auteur, les institutions auxquelles il est affilié lors de la publication, le nombre de co-auteurs, le nombre de pays impliqués dans la publication, la liste de ces pays, la liste de l'ensemble des institutions partenaires, ainsi que deux dictionnaires associant respectivement chaque co-auteur à son institution et chaque institution à son pays.

Surtout, il permet d'identifier l'ensemble des co-auteurs, les institutions auxquelles ils sont affiliés au moment précis de la publication et les pays associés à ces institutions. Ces informations rendent possible l'identification fine des collaborations internationales, tant en termes de volume que de diversité géographique et institutionnelle.

La collecte complète des données a nécessité plusieurs heures de calcul pour l'ensemble des institutions, en raison du volume important de publications et des contraintes imposées par l'API d'OpenAlex. Une première phase de collecte a mis en évidence l'absence de certaines variables essentielles à l'analyse, en particulier le pays d'appartenance des institutions partenaires et l'affiliation institutionnelle précise des co-auteurs pour chaque publication. La procédure de collecte a donc été entièrement relancée afin d'intégrer systématiquement ces informations. Une fois l'ensemble des données collectées, les fichiers produits pour chaque institution ont été concaténés afin de constituer un jeu de données unique. Les observations comportant des erreurs manifestes ont été supprimées.

Dans un second temps, à partir de ce jeu de données concaténé, nous avons constitué un échantillon de chercheurs américains afin de pouvoir comparer leurs trajectoires de collaboration à celles des chercheurs français. Pour ce faire, nous avons récupéré, pour chaque publication associée aux chercheurs français, la liste complète des co-auteurs. Une nationalité a été attribuée à chacun de ces co-auteurs en déterminant le pays dans lequel l'auteur a été le plus fréquemment affilié au cours de sa carrière. Concrètement, pour chaque auteur, nous avons reconstitué la liste des institutions auxquelles il a été affilié, associé à chacune de ces institutions son pays d'appartenance, puis identifié le pays majoritaire parmi ces affiliations institutionnelles.

Sur cette base, nous avons conservé uniquement les auteurs américains, ce qui représente un ensemble de 255 157 chercheurs. Ce choix méthodologique repose sur l'idée que récupérer les

chercheurs américains directement à partir des co-auteurs des chercheurs français permet de mettre en perspective des chercheurs ayant travaillé sur les mêmes publications, rendant ainsi la comparaison de leurs trajectoires de collaboration plus pertinente. Il s'agit donc d'une population de chercheurs américains effectivement connectés aux chercheurs français par des collaborations observées et non d'un échantillon aléatoire indépendant.

À partir de cette population, nous avons mis en œuvre un script similaire à celui utilisé pour les chercheurs français, afin de récupérer pour les chercheurs américains exactement les mêmes métadonnées. Initialement, nous avions envisagé de collecter les données pour 86 380 chercheurs américains, ce nombre ayant été choisi afin de correspondre approximativement au nombre de chercheurs français présents dans notre base de données et d'augmenter la probabilité d'identifier des collaborations effectives avec la France. Toutefois, cette collecte s'est avérée trop coûteuse en ressources de calcul. Par conséquent, elle a finalement été réalisée pour 55 000 chercheurs américains.

Pour chacun des chercheurs américains retenus, le script récupère l'intégralité de leurs publications référencées dans OpenAlex. Chaque ligne du jeu de données correspond à une publication associée à l'auteur considéré et contient exactement les mêmes métadonnées que celles collectées pour les chercheurs français.

L'ensemble de ces données permet de construire un cadre d'analyse homogène pour les chercheurs français et américains et de comparer leurs trajectoires de collaboration internationale en termes de volume et de diversité, en cohérence avec les hypothèses formulées. Effectivement, afin de répondre aux hypothèses 2 et 3, nous travaillons uniquement sur un échantillon de chercheurs américains ayant travaillé sur les mêmes papiers que les chercheurs français dont nous avons des données. Ce sont donc tous des chercheurs ayant collaboré avec la France : cela ne pose pas de problème de comparaison entre la France et les États-Unis puisque dans les deux cas, le groupe de chercheurs étudié ne fait pas l'objet d'une comparaison avec des chercheurs n'ayant jamais de collaboration franco-états-unienne. De plus, puisqu'ils ont tous travaillé sur des papiers en commun, on pourra observer l'effet d'une même collaboration franco-américaine sur les trajectoires des différents chercheurs.

IV. Exploration des données

A. Description des données des tests statistiques

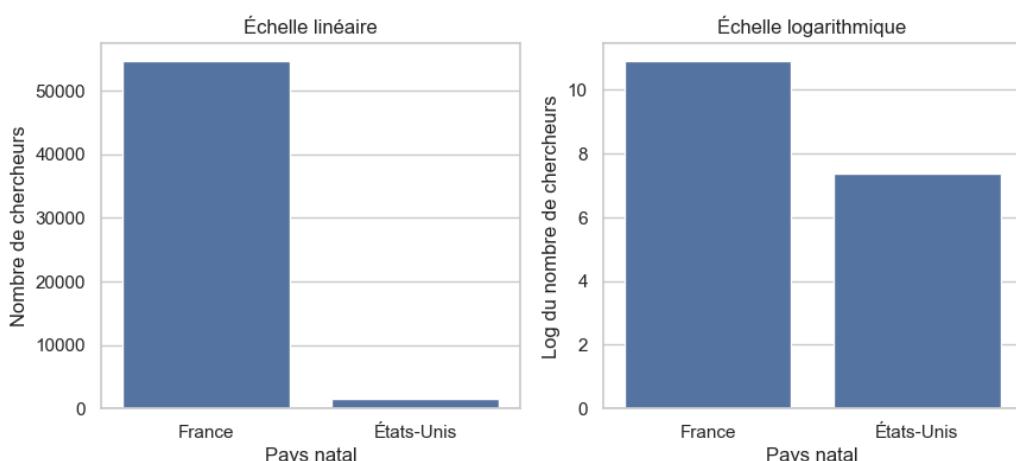


Figure 1. Distribution des chercheurs en fonction du pays d'origine (France ou États-Unis).

Les graphiques de la **Figure 1** présentent la distribution des chercheurs en fonction de leur pays d'origine (France ou États-Unis). On constate une très large prédominance des chercheurs français par rapport aux chercheurs états-uniens dans l'échantillon. Cela est dû à la manière dont nous avons construit nos données.

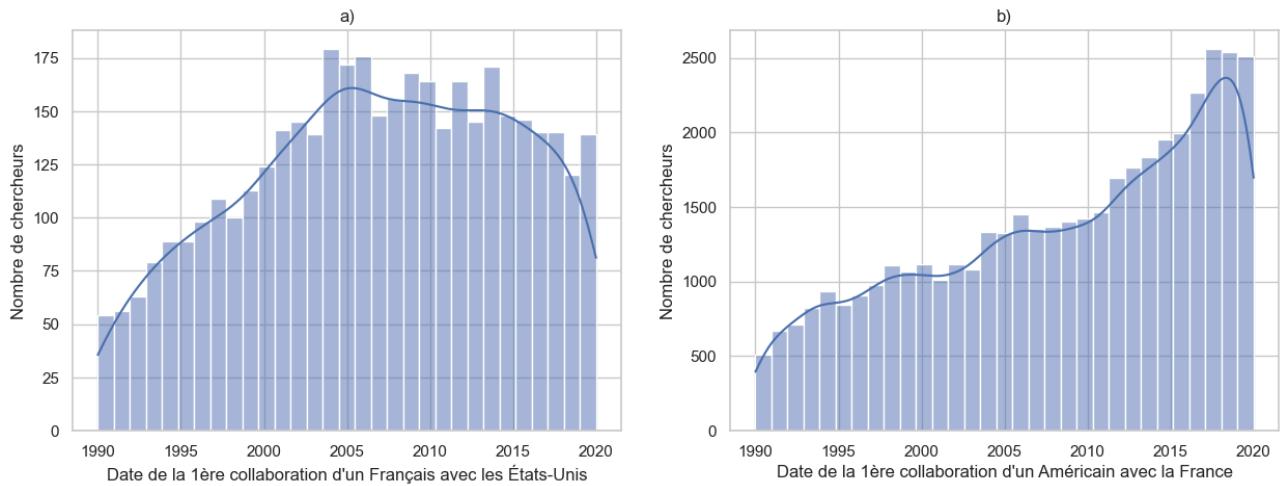


Figure 2. Dates de la première collaboration entre les États-Unis et la France.

Les deux histogrammes de la **Figure 2** illustrent la distribution temporelle des premières collaborations transatlantiques. Le graphique 2.a (première collaboration d'un Français avec les États-Unis) montre une distribution relativement stable au fil du temps, suggérant un intérêt constant des chercheurs français pour la collaboration avec les États-Unis depuis le début des années 1990. En revanche, le graphique 2.b (première collaboration d'un Américain avec la France) révèle une tendance à l'augmentation de la fréquence des premières collaborations avec la France, indiquant une popularité croissante de la France en tant que partenaire de collaboration pour les chercheurs américains sur la période étudiée.

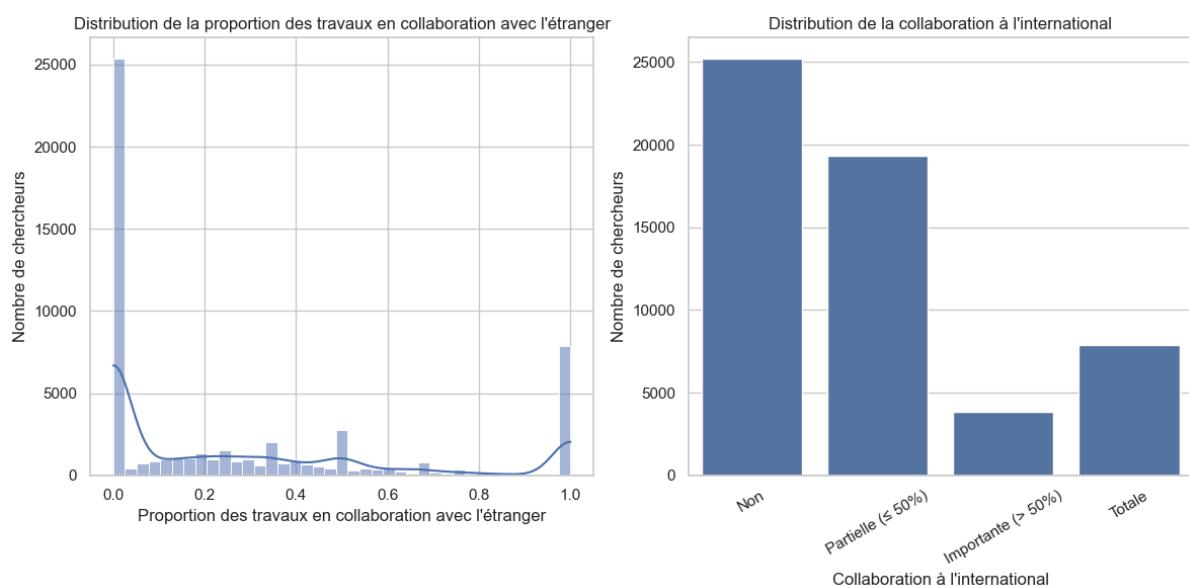


Figure 3. Proportion des travaux en collaboration avec l'étranger.

Enfin les graphiques de la **Figure 3** illustrent deux visualisations de la proportion de publications réalisées en collaboration avec des entités étrangères ("ratio_foreign_collab"). L'histogramme de gauche met en évidence une distribution bimodale avec des pics prononcés aux valeurs de 0 (aucune collaboration étrangère) et 1 (toutes les collaborations sont étrangères), ainsi qu'une concentration de chercheurs ayant une faible proportion de collaborations étrangères (entre 0 et 0.5). Le graphique de droite, qui catégorise cette proportion, confirme ces observations. Il est notable qu'un nombre significatif de chercheurs collaborent exclusivement à l'étranger (catégorie "Totale"). Cette distribution suggère une hétérogénéité des comportements de collaboration, avec des chercheurs soit fortement internationalisés, soit très peu, et une fraction non négligeable ayant une collaboration internationale partielle.

B. Analyse exploratoire portant sur les chercheurs français

Pour poursuivre l'analyse, nous nous intéressons aux chercheurs français. Nous exploitons ici un jeu de données construit au niveau des articles, où chaque ligne correspond à une publication. Cette partie propose une analyse exploratoire centrée sur les trajectoires de publication des chercheurs français et leurs caractéristiques générales. L'objectif est de mettre en évidence les principales régularités observables dans les données, afin de fournir un premier éclairage et de préparer les analyses statistiques plus approfondies présentées par la suite.

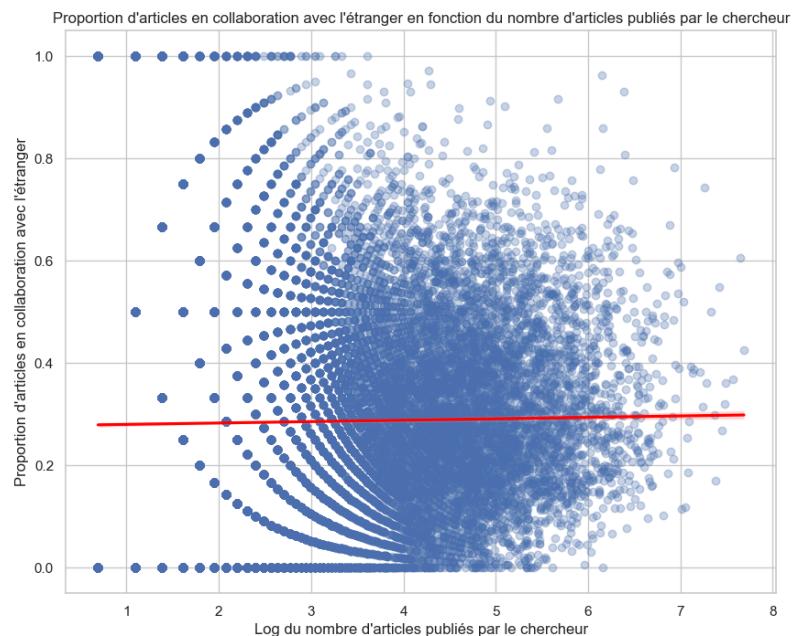


Figure 4. Graphique bivarié de la proportion d'articles avec l'étranger en fonction du nombre d'articles publiés par le chercheur français.

Le graphique bivarié ci-contre représente la relation entre le logarithme du nombre total de publications d'un chercheur français et la proportion de ses articles réalisés en collaboration avec l'étranger. Le nuage de points est très dense aux faibles valeurs de $\text{Log}(1 + \text{total_pubs})$, suggérant qu'une grande majorité des chercheurs français ont un nombre de publications relativement bas. Pour ces chercheurs, la proportion d'articles réalisés avec l'étranger varie sur l'ensemble de l'intervalle $[0, 1]$. À mesure que le nombre de publications augmente (valeurs plus élevées sur l'axe des x), le nuage de points devient plus dispersé et moins dense, indiquant moins de chercheurs avec un grand nombre de

publications. On observe une concentration visible des points au taux de collaboration étrangère : 0 et au taux de collaboration étrangère : 1, ce qui est cohérent avec l'observation des graphiques univariés où une partie des chercheurs n'a aucune collaboration étrangère ou, au contraire, ne collabore qu'avec l'étranger. La pente de la droite de régression linéaire est extrêmement faible. Sa trajectoire n'indique pas une tendance claire qui lierait de manière significative le nombre de publications à la proportion de collaborations étrangères. Cela renforce l'idée qu'être un chercheur très prolifique ne se traduit pas nécessairement par une plus grande proportion de collaborations internationales, et vice-versa. Cela indique que d'autres facteurs, non représentés dans ce graphique bivarié, sont probablement plus déterminants dans la propension à collaborer à l'international ou dans le volume de publications des chercheurs français.

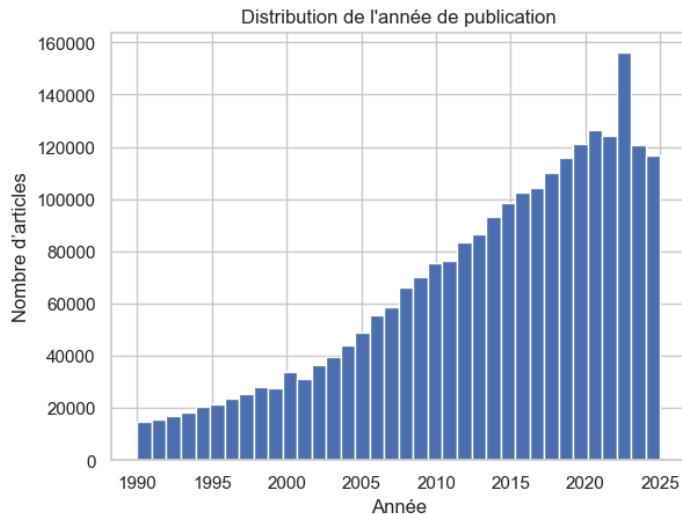
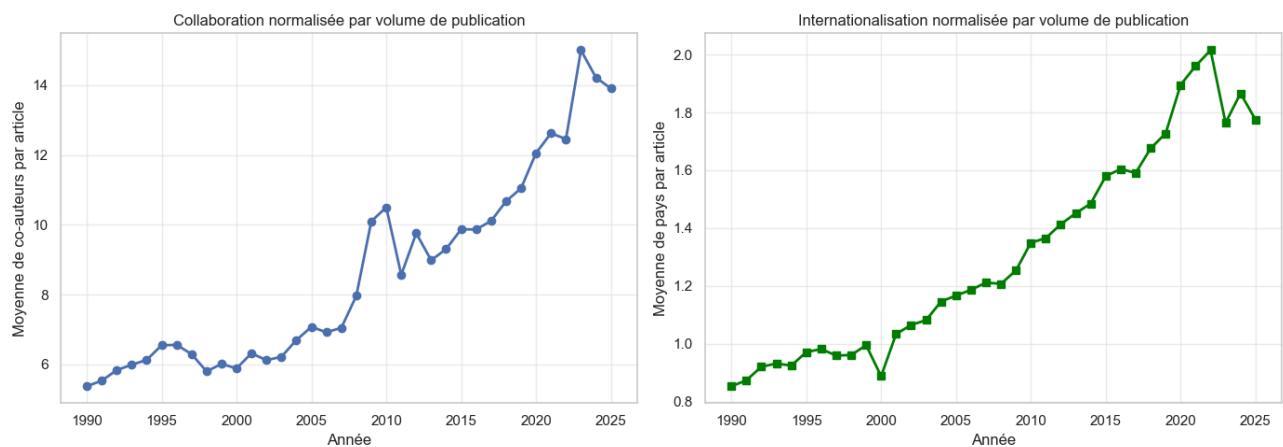


Figure 5. Distribution temporelle et collaboration scientifique.



Le premier graphique de la **figure 5** illustrant la distribution de l'année de publication des articles des chercheurs français montre une tendance claire à l'augmentation du nombre d'articles au fil du temps. Il y a une augmentation progressive du nombre de publications de 1990 jusqu'aux années les plus récentes, ce qui est cohérent avec l'expansion générale de la recherche scientifique et l'accroissement des infrastructures de publication. De plus, le nombre moyen de coauteurs par article augmente au cours du temps. On constate également une internationalisation croissante de la rédaction des articles.

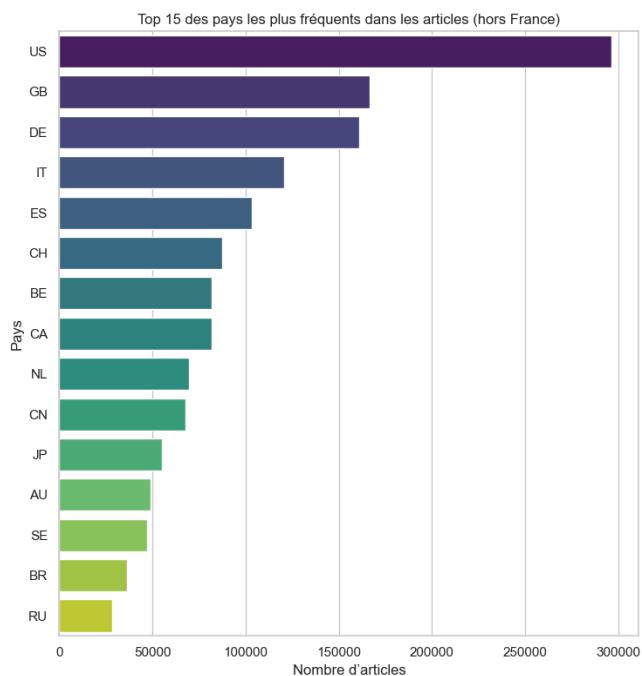
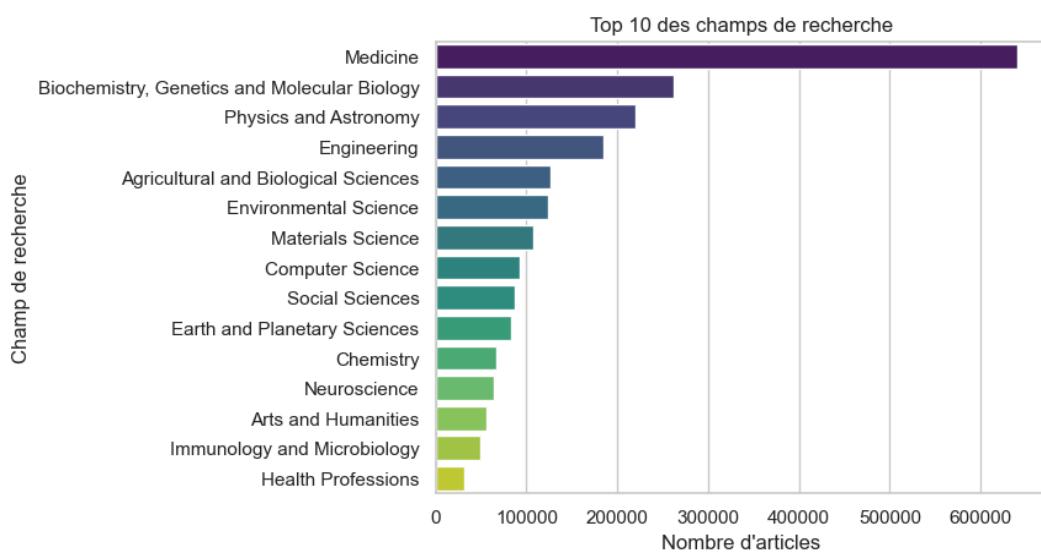


Figure 6. Top 15 des pays les plus fréquents dans les articles (hors France).

Nous avons ensuite regardé la distribution des pays de collaboration des chercheurs français. Le nombre d'articles avec une affiliation française dépasse largement tous les autres pays, soulignant l'ancrage national de la recherche étudiée. Les États-Unis (US) apparaissent comme le collaborateur le plus fréquent, suivis par l'ensemble de ses pays voisins d'Europe de l'Ouest tels que l'Allemagne (DE), le Royaume-Uni (GB), l'Italie (IT) et l'Espagne (ES), ainsi que la Suisse (CH).



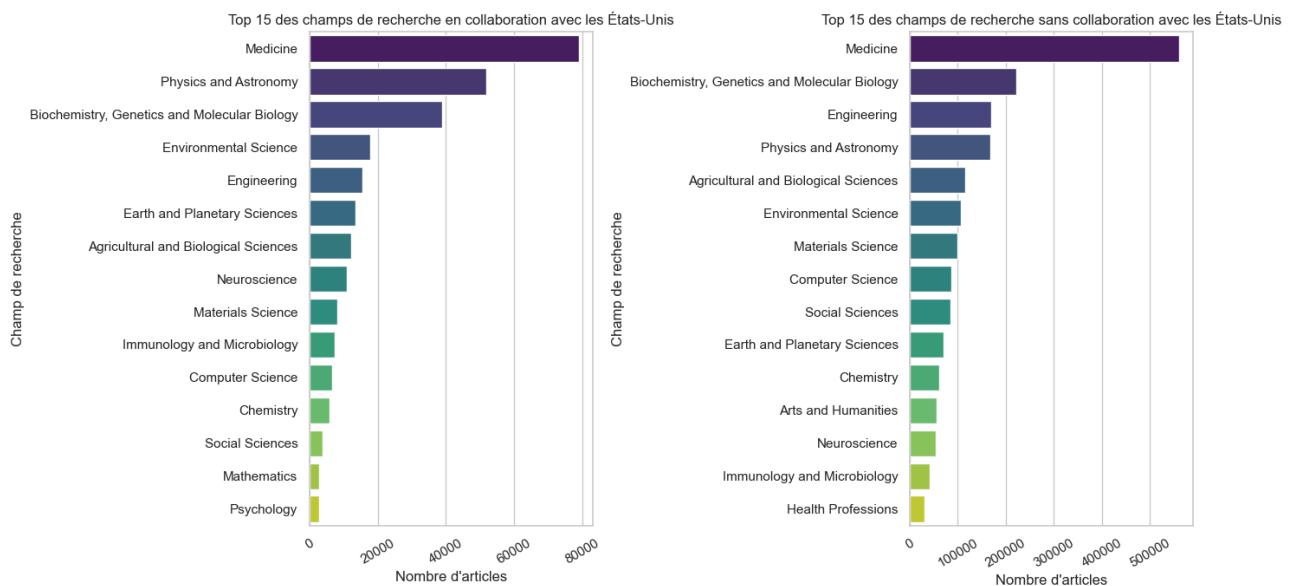


Figure 7. Distribution des champs de recherche.

Le graphique des top 10 des champs de recherche des chercheurs français révèle une prédominance de la recherche dans les sciences dures telles que l'ingénierie, la physique, la chimie, l'informatique et la science des matériaux, aux côtés des sciences sociales. Le second ensemble de graphiques comparant les champs de recherche avec et sans collaboration américaine, apporte des nuances. Les sciences des matériaux, les mathématiques et la psychologie sont plus visibles dans le top 15 lorsque les collaborations avec les États-Unis sont présentes. En revanche, les arts et les sciences humaines, ainsi que les professions de la santé, sont plus représentés dans le top 15 des recherches sans collaboration états-unienne.

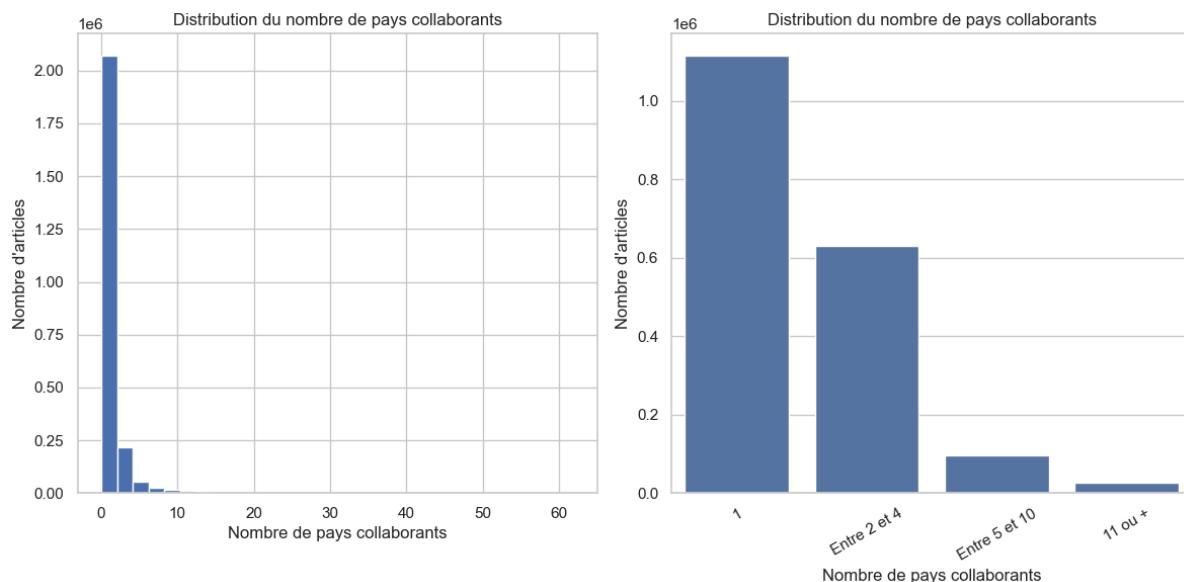


Figure 8. Distribution du nombre de pays collaborants.

Les graphiques de la **Figure 8** illustrent la distribution du nombre de pays impliqués, par article d'un chercheur français. La distribution montre que la grande majorité des articles proviennent d'un seul pays, suivi par un nombre significatif d'articles impliquant 2 à 4 pays. Au-delà, la fréquence diminue rapidement, avec un nombre très faible d'articles impliquant plus de 10 pays. Le premier histogramme

montre une forte concentration sur le « 1 » pays, puis une décroissance rapide, indiquant que la majorité de la production est soit nationale, soit limitée à un petit nombre de collaborations internationales. Ceci suggère que, malgré une tendance à l'internationalisation, une part substantielle de la recherche française se fait dans un cadre national ou avec un petit cercle de partenaires étrangers.

V. Préparation des données pour les tests

Une fois les données françaises et américaines intégralement récupérées et nettoyées, nous avons procédé à la création des mesures nécessaires pour effectuer les tests. Pour ce faire, nous avons construit un nouveau jeu de données structuré de telle sorte qu'une ligne correspond à un auteur. Ce format permet de résumer, pour chaque chercheur, ses trajectoires de collaboration avant et après une première collaboration franco-américaine, et de calculer des indicateurs reflétant à la fois l'intensité et la diversité de ses collaborations internationales.

Parmi les variables principales créées dans ce jeu de données, on retrouve tout d'abord “home_country”, qui indique le pays natal ou principal du chercheur. La variable “first_year_collab” correspond à l'année de la première collaboration entre le chercheur et un partenaire américain (pour les chercheurs français) ou avec un partenaire français (pour les chercheurs américains), lorsque cette collaboration existe.

Pour mesurer l'intensité et l'étendue des collaborations internationales avant et après cette première collaboration, nous avons créé plusieurs variables. “nb_papers_before” correspond au nombre de pays avec lesquels le chercheur a collaboré avant la première collaboration franco-américaine, et “nb_papers_after” au nombre de pays avec lesquels il a collaboré après cette première collaboration. De même, “count_foreign_before” indique le nombre d'institutions étrangères partenaires avant la première collaboration franco-américaine, et “count_foreign_after” le nombre d'institutions étrangères partenaires après cette collaboration. Ces variables permettent de quantifier le nombre total de collaborations internationales et leur évolution dans le temps.

Afin d'analyser l'intensité moyenne de la collaboration par publication, nous avons calculé “intensity_before” et “intensity_after”, définies respectivement comme le ratio de collaborations internationales par publication avant et après la première collaboration franco-américaine, c'est-à-dire “count_foreign_before” / “nb_papers_before” et “count_foreign_after” / “nb_papers_after”. Sur cette base, nous avons construit la variable “evolution_ratio”, correspondant au rapport entre “intensity_after” et “intensity_before”. Cette mesure indique si un chercheur tend à collaborer davantage en moyenne après la première collaboration internationale, un ratio supérieur à 1 signifiant un accroissement de l'intensité.

Nous avons également créé des mesures relatives à la diversification géographique des collaborations. Les variables “geo_intensity_before” et “geo_intensity_after” correspondent au nombre moyen de pays partenaires par publication, avant et après la première collaboration franco-américaine, calculées respectivement comme “nb_countries_before” / “nb_papers_before” et “nb_countries_after” / “nb_papers_after”. À partir de ces deux indicateurs, nous avons défini “geo_diversification_shift”, qui représente le ratio entre “geo_intensity_after” et “geo_intensity_before”. Comme pour le ratio d'intensité, un chiffre supérieur à 1 indique que le chercheur collabore en moyenne avec un plus grand nombre de pays après sa première collaboration franco-américaine.

Dans un souci de contrôler l'effet potentiel de collaborer avec des institutions prestigieuses et de réduire les biais confondants, dans les hypothèses 2 et 3, nous avons intégré les classements de Shanghai des universités mondiales pour la période 2003-2020, en considérant, au minimum, les 100 premières institutions de chacune des années. Pour chaque publication, nous avons identifié le meilleur classement de Shanghai parmi les universités partenaires, que nous avons enregistré dans la variable “best_rank_partners”. À partir de cette information, nous avons construit deux variables supplémentaires pour chaque chercheur : “best_partner_shanghai_before”, correspondant au meilleur classement de Shanghai parmi les institutions partenaires avant la première collaboration franco-américaine, et “best_partner_shanghai_after”, correspondant au meilleur classement des institutions partenaires après cette collaboration. Pour les chercheurs américains, les définitions sont symétriques : il s'agit du meilleur classement des universités partenaires avant et après leur première collaboration avec un chercheur français.

Une fois toutes ces mesures calculées et intégrées dans le jeu de données, ce dernier était prêt à être utilisé pour les analyses statistiques.

VI. Tests statistiques

A. H1 : Effet de la collaboration franco-américaine sur l'internationalisation des travaux

Pour mesurer l'impact d'une collaboration franco-américaine pour un chercheur français sur la proportion internationale de ses publications, nous avons visualisé le diagramme en boîte suivante. La figure ci-dessous compare les chercheurs français selon qu'ils ont ou non collaboré avec les États-Unis.

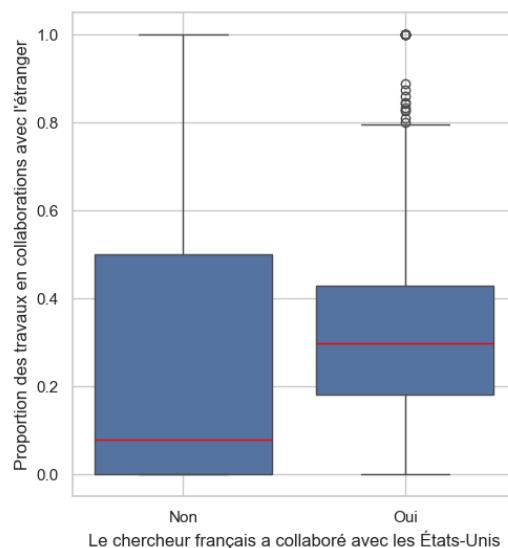


Figure 9. Proportion des travaux d'un chercheur français à l'étranger en fonction de la collaboration s'il a collaboré avec les Etats-Unis .

Note : Chaque boîte représente l'étendue interquartile de la proportion de travaux internationaux, la ligne rouge indique la médiane.

Le graphique met en évidence plusieurs tendances : les chercheurs français peuvent collaborer à l'international même sans partenariat avec les États-Unis. La proportion de travaux internationaux est nettement plus élevée pour les chercheurs français ayant collaboré avec les États-Unis, ce qui suggère un rôle important de la première collaboration franco-américaine dans l'internationalisation des publications.

Pour quantifier cette relation, nous avons réalisé un test statistique comparant la proportion de publications internationales entre les Français ayant collaboré avec les États-Unis et ceux ne l'ayant pas fait. Les résultats indiquent que les chercheurs français ayant collaboré avec les États-Unis présentent une proportion de publications internationales significativement plus élevée que ceux n'ayant jamais collaboré avec les États-Unis ($t = 4.52$, $p < 0.001$).

Ces résultats confirment que la première collaboration internationale franco-américaine est associée à une augmentation significative de la proportion de publications en collaboration avec des institutions étrangères pour les chercheurs français. Le *boxplot* permet de visualiser cette tendance, illustrant l'accroissement de la médiane et de la dispersion de la proportion de collaborations internationales dans le groupe ayant participé à une collaboration franco-américaine.

B. H2 : Effet de la première collaboration franco-américaine sur l'évolution du nombre moyen d'institutions partenaires par publication

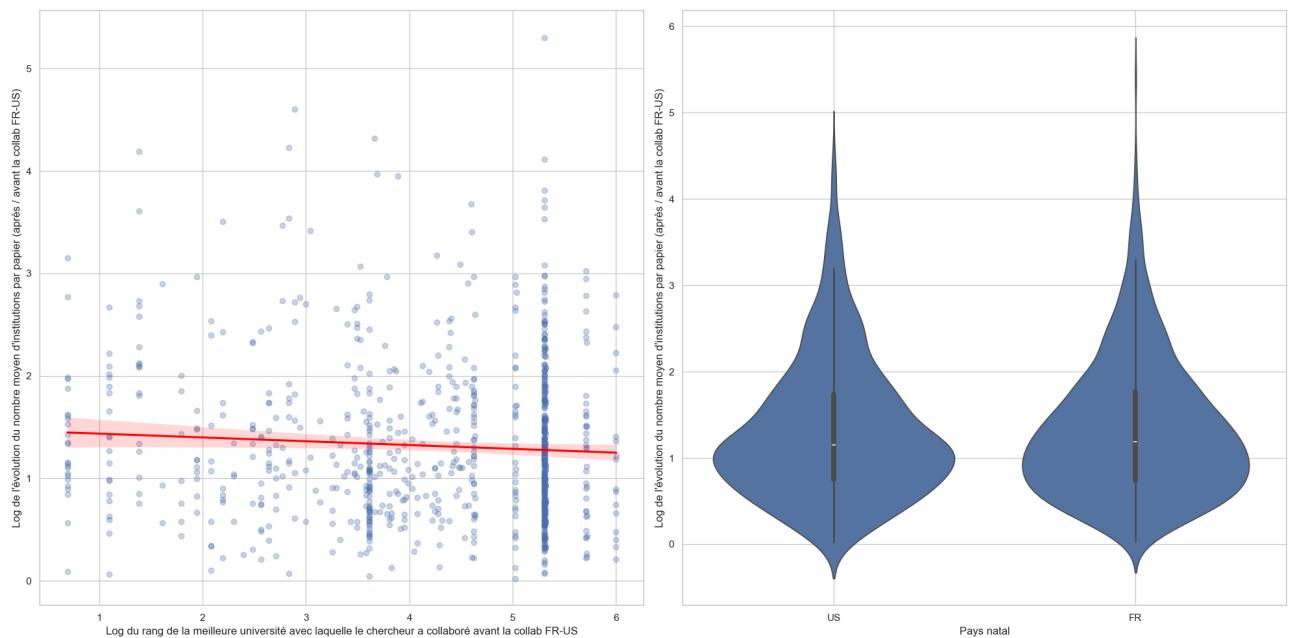


Figure 10. Évolution du nombre moyen d'institutions partenaires par publication après la première collaboration franco-américaine.

Pour évaluer l'impact de la première collaboration franco-américaine sur l'évolution du nombre moyen d'institutions partenaires par publication, nous avons produit deux visualisations complémentaires. Le premier graphique de la **Figure 10** représente la relation entre le rang de la meilleure université partenaire avant la première collaboration franco-états-unienne (échelle logarithmique) et le ratio d'évolution du nombre moyen d'institutions partenaires par papier (échelle logarithmique). Le second graphique compare la distribution de ce ratio d'évolution entre les chercheurs français et américains (échelle logarithmique) à l'aide d'un diagramme violon.

L'analyse graphique descriptive (analyse bivariée) met en évidence une absence de corrélation linéaire apparente, ce que le modèle multivarié viendra confirmer par la suite.

Le graphique de régression met en évidence une droite quasi horizontale, cela suggère que le rang de la meilleure université partenaire avant la première collaboration franco-américaine n'est pas associé de manière notable à l'évolution du nombre moyen d'institutions partenaires par publication. Autrement dit, le prestige des partenaires antérieurs ne semble pas conditionner l'intensification ultérieure des collaborations institutionnelles.

Le diagramme violon montre que, pour les deux nationalités, la distribution du logarithme du ratio d'évolution du nombre moyen d'institutions partenaires par publication est majoritairement située au-dessus de zéro. Cela indique que, pour une large part des chercheurs français comme américains, le nombre moyen d'institutions partenaires par papier tend à augmenter après la première collaboration franco-américaine. Un logarithme du ratio d'évolution supérieur à zéro correspond en effet à une augmentation du nombre moyen d'institutions partenaires par publication après la collaboration franco-états-unienne, tandis qu'une valeur négative indique une diminution. La dispersion observée révèle toutefois une hétérogénéité importante des trajectoires individuelles, certains chercheurs connaissant une augmentation marquée de leurs collaborations institutionnelles, tandis que d'autres présentent une évolution plus modérée.

| Colonne | Coefficient | P-value |
|------------------------------|-------------|---------|
| home_country[T.US] | 0.4623 | 0.558 |
| best_shanghai_partner_before | -0.2656 | 0.503 |

Table 1. Résultats de la régression linéaire multivariée d'équation :

$$\text{evolution_ratio}_i = \beta_0 + \beta_1 \text{home_country}_i + \beta_2 \text{best_partner_shanghai_before}_i.$$

$$R^2 = 0.002.$$

Pour quantifier ces relations, nous avons estimé un modèle de régression linéaire incluant simultanément le rang de la meilleure université partenaire avant la collaboration franco-états-unienne et la nationalité du chercheur comme variables explicatives. Les résultats montrent qu'aucune de ces variables n'est statistiquement significative ($\beta = 0.4623$, $p = 0.558$ pour la nationalité ; $\beta = -0.2656$, $p = 0.503$ pour le rang de l'université partenaire), indiquant que ni le prestige des partenaires antérieurs ni la nationalité du chercheur n'expliquent de manière robuste l'évolution du nombre moyen d'institutions partenaires par publication.

Ces résultats suggèrent que, bien que la première collaboration franco-américaine soit associée à une tendance générale à l'augmentation du nombre moyen d'institutions partenaires par papier pour les chercheurs français comme américains, cette évolution ne peut pas être attribuée de manière significative à la nationalité ou au prestige des collaborations antérieures. Les graphiques permettent ainsi de visualiser des dynamiques individuelles d'intensification des collaborations internationales, sans mettre en évidence d'effet différencié ou structurel entre les deux groupes.

C. H3 : Diversification géographique des collaborations après la première collaboration franco-américaine

Afin d'évaluer l'évolution de la diversité géographique des collaborations internationales après une première collaboration franco-américaine, nous nous intéressons au changement du nombre moyen de pays partenaires par publication, mesuré par le ratio entre la période postérieure et antérieure à cette première collaboration franco-américaine. Deux visualisations complémentaires ont été produites pour analyser cette évolution.

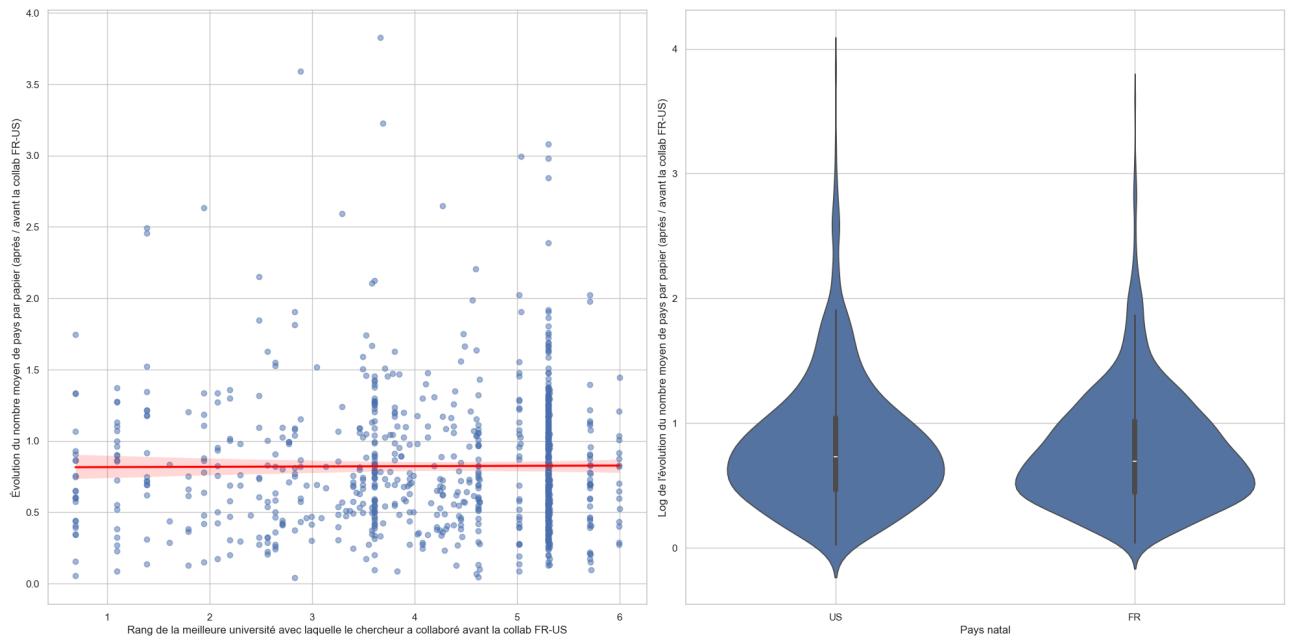


Figure 11. Évolution de la diversification géographique des collaborations internationales après la première collaboration franco-américaine.

La **Figure 11** présente, à gauche, la relation entre le rang de la meilleure université partenaire avant la première collaboration franco-américaine (en échelle logarithmique) et le logarithme du ratio d'évolution du nombre moyen de pays partenaires par publication. À droite, un diagramme violon compare la distribution de ce log du ratio d'évolution selon la nationalité du chercheur (France / États-Unis).

L'analyse graphique descriptive (analyse bivariée) met en évidence une absence de corrélation linéaire apparente ici aussi, ce que notre deuxième modèle multivarié viendra confirmer par la suite.

L'examen du graphique de gauche montre que la droite de régression est quasiment horizontale, indiquant une absence de relation marquée entre le prestige académique des partenaires antérieurs à la collaboration franco-américaine et l'évolution ultérieure de la diversification géographique des collaborations. Autrement dit, le rang de la meilleure université partenaire avant la collaboration franco-américaine ne semble pas associé à une augmentation ou à une diminution systématique du nombre moyen de pays partenaires par publication.

Le diagramme violon suggère que, pour les deux nationalités, la distribution du logarithme du ratio d'évolution est globalement centrée autour de valeurs positives, ce qui indique qu'en moyenne, les chercheurs tendent à collaborer avec un nombre légèrement plus élevé de pays par publication après leur première collaboration franco-états-unienne. Toutefois, les distributions observées pour les

chercheurs français et américains sont très proches, tant en termes de position centrale que de dispersion, ce qui ne permet pas d'identifier visuellement une différence nette entre les deux groupes.

| Colonne | Coefficient | P-value |
|------------------------------|-------------|---------|
| home_country[T.US] | 0.1260 | 0.549 |
| best_shanghai_partner_before | -0.0768 | 0.467 |

Table 2. Résultats de la régression linéaire multivariée d'équation :

$$\text{geo_diversification_shift}_i = \beta_0 + \beta_1 \text{home_country}_i + \beta_2 \text{best_partner_shanghai_before}_i.$$

$$R^2 = 0.002.$$

Ces observations graphiques sont confirmées par l'estimation d'un modèle de régression linéaire intégrant simultanément la nationalité du chercheur et le rang de la meilleure université partenaire avant la collaboration franco-états-unienne. Aucun des deux effets n'apparaît statistiquement significatif ($\beta = 0.1260$, $p = 0.549$ pour la nationalité ; $\beta = -0.0768$, $p = 0.467$ pour le rang de l'université partenaire). Ces résultats indiquent qu'il n'existe pas de relation statistiquement robuste entre la diversification géographique des collaborations après la première collaboration franco-américaine et, d'une part, le prestige des partenaires antérieurs, ni, d'autre part, la nationalité du chercheur.

Ainsi, bien que les chercheurs français comme américains tendent individuellement à accroître légèrement la diversité géographique de leurs collaborations après une première collaboration franco-états-unienne, cette évolution ne diffère pas significativement entre les deux groupes. L'hypothèse selon laquelle les chercheurs français diversifient davantage leurs collaborations internationales que les chercheurs américains à la suite d'une première collaboration franco-américaine n'est donc pas confirmée par les analyses empiriques.

VII. Conclusion

Les résultats montrent que la collaboration universitaire franco-américaine est associée à une intensification significative de l'internationalisation des publications pour les chercheurs français, mais sans effet différencié robuste en faveur des chercheurs français par rapport aux chercheurs américains en termes d'évolution ou de diversification ultérieure de leurs collaborations internationales. Cela suggère que si cette collaboration constitue un levier de rayonnement scientifique, ses effets apparaissent globalement symétriques entre chercheurs français et américains.