

Demande de savoir et reconfigurations du champ scientifique : les effets des épidémies de corononavirus sur les dynamiques de réussite des chercheurs

Maxime Chabriel^{*,1} et Mathis Sansu^{†,1}

¹ENSAE Paris, France

11 mai 2022

Résumé

La pandémie du COVID-19 a amené des évolutions du champ scientifique par un changement d'orientation des sujets traités et par l'organisation du travail, des habitudes et des normes du monde de la recherche scientifique. Cet article interroge comment les crises épidémiques dues à des coronavirus amènent des restructurations de ce champ de recherche. Nous proposons un modèle théorique où chaque auteur dispose de capitaux, endogène au sous-champ et donc plus directement mobilisables, et exogène traduisant la réussite dans le champ scientifique hors du domaine de recherche des coronavirus. Usant de données extraites de la base Scopus, nous identifions que la réussite future dans le sous-champ des coronavirus est en partie déterminée par une participation antérieure à ce domaine de recherche, et ce davantage qu'un investissement dans un autre champ de recherche. Les moments où la demande de savoir est intense - induisant un nombre élevé de primo-auteurs dans le domaine - structurent une ascendance hiérarchique scientifique des auteurs déjà installés. C'est particulièrement le cas à la suite de la crise du COVID-19 qui semble restructurer le champ fortement autour des auteurs insérés dans le domaine peu avant l'explosion pandémique.

Remerciements

Nous tenons à remercier sincèrement Ivaylo Petev, qui nous a accompagné et conseillé tout au long de la réalisation de ce travail. Son expérience et ses suggestions nous ont été d'une grande aide pour enrichir cette recherche. Il serait ingrat de ne pas remercier chaleureusement Julie et Cécile, nos camarades qui ont dû supporter notre duo insupportable, et qui ont de plus largement contribué à rendre ce travail plus solide, poussé et achevé par leurs remarques et questions, avec une attitude toujours bienveillante.

1 Introduction

Malgré le premier recensement d'une infection par un coronavirus chez l'homme en 1965, le domaine de la recherche sur cette famille de pathogènes n'est devenu un sujet scientifique de premier plan qu'à partir des épidémies dites à syndrome respiratoire aigu sévère (SRAS). Les épidémies du SARS-CoV (2002-2004), du MERS-CoV (2012-2014), et du SARS-CoV-2 (2020-2022) ont tour à tour augmenté la demande de connaissances sur ce type de virus et motivé davantage d'acteurs du monde de la recherche - scientifiques, laboratoires, journaux de publication - à se pencher sur leur étude. Du point de vue de la sociologie des sciences, ces épidémies possèdent un intérêt méthodologique particulier : ce sont des événements soudains, non anticipés, et exogènes au monde de la recherche. Ce sont donc des expériences naturelles de l'impact sur la

communauté scientifique d'un choc positif de demande de savoir sur un domaine de recherche relativement restreint. De par ces caractéristiques originales, le champ de la recherche sur les coronavirus peut alors nous permettre de comprendre comment l'évolution de l'attractivité d'un champ de recherche a un impact sur les parcours professionnels de ses individus. En particulier, on peut s'interroger sur les caractéristiques des nouveaux acteurs *i.e.* les scientifiques qui décident de publier pour la première fois un article de recherche sur le thème des coronavirus. Mais aussi, ce que deviennent les acteurs préalablement installés dans le champ : sont-ils davantage sollicités pour leur expertise, ou perdent-ils au contraire en légitimité au profit de ces nouveaux entrants ? Si l'on assiste à un effet d'éviction - c'est-à-dire une diminution relative de l'importance au sein du champ pour les chercheurs s'y trouvant préalablement -, quels sont les critères qui favorisent l'installation des

^{*}maxime.chabriel@ensae.fr

[†]mathis.sansu@ensae.fr

nouveaux entrants et le départ des sortants ? Nous tentons de répondre à ces questions en recourant théoriquement aux caractéristiques structurantes du champ scientifique et grâce à l'application de modèles de régression aux moindres carrés ordinaires (MCO) sur des observations construites à partir de la base de données Scopus.

Cet article est structuré de la façon suivante : la [seconde section](#) effectue une revue de la littérature du sujet traité. La [troisième partie](#) définit nos hypothèses de recherche et propose un modèle théorique pour les vérifier. La [partie suivante](#) présente les données utilisées ainsi que leur procédure d'extraction. La [section 5](#) évoque les méthodes d'analyses empiriques adoptées, avant que la [partie suivante](#) ne s'étende sur les résultats obtenus. La [section 7](#) discute ces derniers tandis que la [section 8](#) conclut.

2 Revue de littérature

L'étude de la production scientifique, des dynamiques des domaines de recherche et des caractéristiques structurantes du milieu de la recherche scientifique ont connu une intensité grandissante avec le développement des données numériques et de leur disponibilité ([Fortunato et al. \(2018\)](#)). La science de la science - *SciSci* - s'attelle notamment à la compréhension de l'effet de la réputation des auteurs sur l'impact des publications scientifiques. [Petersen et al. \(2014\)](#) estiment ainsi que la réputation d'un auteur joue un rôle majeur dans le nombre de citations d'une nouvelle publication scientifique, au moins jusqu'à un certain seuil de citations à partir duquel l'effet de la réputation devient négligeable. D'une certaine façon, le traçage numérique et la quantification de la réussite d'un article - à travers l'automatisation par certaines revues scientifiques de la publication de scores pour chaque article publié quant à leur nombre de citations, le classement par popularité, voire même parfois la création de métriques évaluant la performance de publication des chercheurs eux-mêmes - a donné forme à la systématisation formelle d'un processus que l'anthropologue Bruno Latour mettait déjà en avant dans les années 1970 ([Latour et al. \(2013\)](#)) : le milieu scientifique produit de manière structurante une compétition mesurée à l'aune des critères de publication, où la réussite antérieure a un impact sur la réussite présente et future. La dynamique d'accumulation de prestige au sein du milieu scientifique est donc centrale pour les auteurs participant à un certain domaine de recherche, notamment si l'on prend compte de l'importance que revêtent aujourd'hui les métriques d'impact dans le milieu scientifique. D'autres facteurs tels que le nombre de co-auteurs sur un article sont étudiés en ce sens pour expliquer les différences d'impact entre articles ([Larivière et al. \(2015\)](#)), permettant ainsi d'éviter l'écueil des auto-citations dans les analyses. In-

tégrant ces résultats, nous proposons de conceptualiser ces traits structurants comme des objets sociaux, à la fois construits et perçus. La modélisation des dynamiques d'impacts des productions scientifiques par auteur doit donc se constituer selon nous comme le résultat d'un espace social hiérarchiquement structuré.

Notre travail théorique est une application empirique des concepts développés par [Bourdieu \(1976\)](#), qui conçoit le champ scientifique comme un espace social où les acteurs - scientifiques, laboratoires, *etc.* - sont inégalement dotés en capitaux scientifiques, produisant une hiérarchie structurante du champ. A l'instar de [Jæger et Breen \(2016\)](#) qui développent un modèle empirique de la théorie de reproduction du capital culturel, nous souhaitons opérationnaliser les dynamiques de réussite au sein du champ scientifique.

Plus particulièrement, nous nous focalisons sur le champ de recherche des coronavirus. Ce choix tient à l'existence de chocs exogènes (épidémies et pandémie susmentionnées) qui impactent directement la production scientifique au sein du champ ([Aviv-Reuven et Rosenfeld \(2021\)](#)). De plus, ces chocs redéfinissent la hiérarchie des acteurs selon leurs caractéristiques sociales : à titre d'exemple, [Squazzoni et al. \(2021\)](#) montrent que les femmes furent les plus pénalisées durant la première vague pandémique de 2020, celles-ci soumettant moins de travaux que leurs homologues masculins. C'est particulièrement le cas des chercheuses des cohortes les plus jeunes : le conflit entre le travail à distance et les contraintes familiales pesant plus lourd sur les femmes que les hommes. Dès lors, nous cherchons à rendre compte des processus qui expliquent la "réussite" des auteurs dans le champ, cela directement à partir de leurs capitaux scientifiques. Notre travail contribue à la littérature en étant, à notre connaissance, le premier à intégrer une modélisation de capitaux scientifiques caractérisés par leur endogénéité / exogénéité à un sous-champ scientifique dans l'explication des dynamiques de réussite des chercheurs.

3 Modèle théorique

3.1 Concepts

Nous faisons le choix de délimiter notre période d'étude selon 4 sous-périodes correspondant aux intervalles d'irruption des crises épidémiques et pandémiques liées à des coronavirus : (i) pré-2003, (ii) 2003-2012, (iii) 2013-2019, (iv) 2020-2022. Les trois crises considérées sont celles du SARS-CoV (2002-2004), du MERS-CoV (2012-2014) et du SARS-CoV-2 (2019-2022). Nous inspirant de [Bourdieu \(1976\)](#), nous considérons le champ de la recherche scientifique biomédicale comme notre objet d'étude, à même de subir des évolutions et reconfigurations au cours de la période et à la suite des "chocs" de demande de savoir. Nous défi-

nissons ce champ comme l'ensemble des acteurs - chercheurs - et institutions - laboratoires, journaux scientifiques - participant à la production de savoirs dans le domaine biomédical. Nous le caractérisons comme un champ, car les acteurs poursuivent un objectif commun et spécifique à ce microcosme social - être publié dans des revues scientifiques de prestige, obtenir une place dans un laboratoire de recherche, être cité par ses pairs, *etc.* - en se conformant aux règles propres définies au sein de cet espace. La réussite individuelle, c'est-à-dire le degré de reconnaissance par les pairs de la compétence scientifique, produit une hiérarchie - un auteur est plus ou moins cité, plus ou moins publié, dans des journaux plus ou moins reconnus, relativement à d'autres auteurs. Progresser dans cette hiérarchie dépend en partie de capitaux sociaux, c'est à dire de caractéristiques et d'accomplissements individuels socialement valorisés au sein du champ - l'expertise, l'expérience, le talent, la renommée, la centralité dans un réseau de chercheurs, *etc.*

Symétriquement, nous définissons par la suite le sous-champ de la recherche biomédicale sur les coronavirus comme l'ensemble des acteurs participant à la production de savoirs dans le domaine biomédical. Cette distinction nous apparaît pertinente dans la mesure où l'activité d'un chercheur se limite à un nombre limité de domaines d'études, souvent connectés, et, que ce soit pour des contraintes liées à une limite d'expertise, un manque de connections, ou simplement un manque de reconnaissance de la part des pairs, il lui est difficile d'en sortir. Nous estimons que la hiérarchie dans le sous-champ des coronavirus ainsi présentée est le produit de plusieurs facteurs, qui sont internes ou externes au sous-champ en question. Nous définissons dès lors deux types de capitaux scientifiques : (i) endogène et (ii) exogène. Le premier est interne au sous-champ (d'où le terme d'endogène), car il est le produit de la participation directe d'un individu à la production des connaissances sur les coronavirus. Le second est dit exogène, car il correspond à la contribution et la reconnaissance dans le champ scientifique de la recherche biomédicale dans son entièreté - duquel on exclut le sous-champ spécifique de la recherche sur les coronavirus. Dans l'hypothèse où les deux types de capitaux sont utiles pour pouvoir réussir dans le sous-champ de la recherche sur les coronavirus, le capital endogène traduit une expertise sur le thème précis des coronavirus tandis que le capital exogène s'apparente davantage à un prestige ou une réputation globale qui est transférable d'un champ à l'autre. Nous pouvons dès lors conceptualiser deux idéaux-types de chercheurs contribuant à la recherche sur les coronavirus : l'*outsider* d'une part, un nouvel entrant au sous-champ ne possédant aucun capital endogène au moment t donné ; au contraire de l'*insider*, un auteur inséré dans le sous-champ des co-

ronavirus - et possédant donc un capital endogène au même moment t .

3.2 Hypothèses

1- Quand les *insiders* peuvent jouer de leur capital d'expertise, les *outsiders* ne possèdent que leur capital exogène pour réussir dans leur nouveau champ. Les *insiders* possèdent donc un avantage initial par rapport aux *outsiders*, et l'acquisition de capital endogène par les *outsiders* se réalise progressivement. **H1** : Après le choc épidémique, toutes choses égales par ailleurs, on s'attend donc à davantage de réussite de la part des *insiders* au sein du champ de recherche des coronavirus que des *outsiders*. De plus, l'inégalité entre ces deux types de chercheurs sera décroissante du temps passé par l'*outsider* au sein du champ.

2- Le capital exogène a un impact sur la position hiérarchique dans le champ et contribue à l'insertion d'un chercheur dans le champ de la recherche sur les coronavirus indépendamment de son capital endogène. Les dynamiques de la hiérarchie scientifique relative du champ de recherche des coronavirus à la suite des chocs épidémiques doivent traduire des effets hétérogènes selon le capital exogène de chaque individu au sein du champ plus large de la recherche scientifique. **H2** : La réussite des *outsiders* - être publié, être cité - au sein du champ de recherche des coronavirus, qui par définition ne peut dépendre d'un capital endogène, est déterminée positivement par leur capital exogène.

3- Aussi, le choc de demande n'a pas seulement permis l'entrée d'un grand nombre d'*outsiders* dans le champ des coronavirus - on peut se référer aux [Figures 3 et 4](#) qui montrent l'évolution quantitative du nombre d'auteurs dans le sous-champ des coronavirus et traduisent bien l'entrée d'*outsiders* aux moments du SARS-CoV et COVID-19 notamment -, il a de plus redéfini les frontières et la structure du champ. Par exemple, les journaux scientifiques publient beaucoup plus sur les coronavirus après le choc qu'avant le choc. Les chocs de demande ont par conséquent tendance à agrandir - au moins pour un temps et de façon parfois très considérable - le champ, l'explosion du nombre d'*outsiders* peut potentiellement diminuer l'efficacité du capital endogène à établir une hiérarchie scientifique. **H3** : Après un choc, on s'attend alors à ce que le capital endogène connaisse une diminution de son importance relative par rapport au capital exogène dans la structuration de la hiérarchie au sein du champ.

4 Données

Les données utilisées pour cette étude sont extraites de Scopus¹, une base de données entretenue par l'éditeur de contenus scientifiques *Elsevier*, qui recense des

1. <https://www.scopus.com/>

références de publications scientifiques ainsi que des variables leur étant afférentes (Baas *et al.* (2020)). Les données ont été intégralement collectées dans le courant de mars 2022. Nous reprenons la méthodologie d'extraction de CORD-19 (Covid-19 Open Research Dataset) afin d'identifier les articles scientifiques traitant des coronavirus dans la base Scopus (Lu Wang *et al.* (2020)). Les articles sont sélectionnés à partir d'une liste de mots-clés² présents au sein des titres ou résumés d'introduction des articles. Pour chacune des 4 périodes que nous avons définies plus haut, nous identifions les 15 journaux scientifiques qui ont le plus contribué au champ des coronavirus en termes de nombre d'articles publiés. Etant donné qu'un journal peut être parmi les 15 plus gros contributeurs pour des périodes différentes, nous obtenons *in fine* une liste de 39 journaux distincts (*cf.* Table 1 en Annexes). Une fois ces 39 journaux identifiés, nous extrayons tous les articles se référant aux coronavirus selon la même méthodologie que celle décrite plus haut pour chacun des journaux et sur la totalité de la période d'étude. Cette procédure nous permet de rendre compte des recompositions du champ temporellement.

Ayant extrait tous les articles sur les coronavirus des 39 journaux, nous dressons la liste de tous les auteurs ayant contribué au domaine, pour un total de plus de 500 000 individus. Pour des raisons computationnelles³, cette liste est restreinte selon plusieurs critères. Tout d'abord, nous faisons le choix d'exclure les auteurs ayant moins 2 publications identifiées comme contribuant au domaine de recherche sur les coronavirus (300 000 individus). Ce critère se justifie par la volonté de diminuer le bruit au sein de nos données en faisant en sorte qu'un auteur n'arrive pas "par hasard" dans le domaine d'étude⁴. De plus, comme nous étudions une trajectoire au sein du sous-champ des coronavirus, il y a un obstacle méthodologique évident à traiter les individus n'ayant qu'une unique apparition dans le sous-champ en question. La population de 200 000 individus ainsi retenue est ensuite à nouveau réduite à un échantillon représentatif stratifié⁵, de 10 109 auteurs. La puissance de calcul dégagé par l'échantillonnage nous permet d'extraire les références de tous leurs articles recensés par Scopus (445 522 articles, 438 846 lorsque l'on exclut les articles à plus de 50 auteurs).

Sont construites par la suite des variables sur la carrière de chacun des auteurs de l'échantillon au sein du champ scientifique biomédical et du sous-champ de la recherche sur les coronavirus.

5 Méthodes

5.1 Opérationnalisation

A partir des données extraites de la base Scopus, nous reconstruisons les métriques classiques d'évaluation de la production scientifique : nous mobilisons le nombre de contributions, le nombre de citations, le nombre de coauteurs d'un auteur. Pour chaque année t , nous établissons des métriques liées à l'accumulation de capitaux sur toutes les années précédant t , ainsi que la réussite - l'accumulation de capitaux - lors de l'année t . Scopus ne fournit que le nombre de citations cumulées sur toute la période lors de l'extraction des données (mars 2022), et non pour une année donnée. Il nous est impossible de savoir si un article réalisé dans les années 1990 et abondamment cité l'a été au cours des années 1990 ou à la suite de la crise du COVID-19 par exemple. Cela peut poser problème dans la mesure où les articles récents peuvent avoir tendance à être davantage crédibles et avoir une meilleure visibilité que des articles plus anciens. Cependant, comme nous effectuons l'analyse de l'évolution des relations causales entre capital scientifique et réussite d'une année à l'autre, nous effectuons une succession de modèles économétriques qui, individuellement, n'introduisent pas de dimension temporelle. Ainsi, nous sommes en mesure de faire l'hypothèse que les effets d'ancienneté sur le nombre de citations que possèdent un article sont homogènes sur tous les articles d'une même année t , rendant la comparaison possibles entre ceux-ci pour t fixé.

Pour construire et différencier le capital exogène de l'endogène, ainsi que la réussite au sein du sous-champ de la recherche des coronavirus de celle au sein du champ global de la recherche biomédicale (duquel est exclu le sous-champ des coronavirus), nous recalculons les variables de chaque auteur sur leur production scientifique liée à leur recherche sur les coronavirus (toujours en distinguant les articles à partir de la méthodologie CORD-19) et ensuite sur leurs autres pro-

2. "COVID-19", "Coronavirus", "Corona virus", "2019-nCoV", "SARS-CoV", "MERS-CoV", "Severe Acute Respiratory Syndrome", "Middle East Respiratory Syndrome"

3. Analyser 500 000 auteurs sur une période de 20 ans reviendrait à étudier une base de deux millions d'observations. L'aspiration automatique des données de cette masse d'auteurs prendrait de nombreux jours. De plus, les variables de type réseau nécessitent des algorithmes à haute complexité computationnelle en fonction du nombre d'individus traité.

4. En particulier, certains articles recensent un grand nombre de co-auteurs, pouvant aller largement au-delà de la dizaine. 1.5% des articles recensés sont signés par 50 auteurs ou plus. Il nous est difficile d'établir que chacun des individus recensés dans cette liste sont des membre actifs du sous-champ de la recherche sur les coronavirus. Par conséquent, nous avons fait le choix de supprimer de notre base tous les articles recensés avec plus de 50 auteurs.

5. L'échantillon est stratifié par les différentes périodes d'activité de l'auteur dans le champ des coronavirus, le nombre d'articles sur les coronavirus publiés sur l'ensemble de notre période d'étude, le nombre de citations que ses articles ont obtenu sur notre période d'étude, le nombre d'articles publiés dans *Plos One* - le journal identifié comme le plus généraliste, c'est-à-dire avec le plus gros volume de publications dans le sous-champ lors de notre période d'étude -, le nombre d'articles publiés dans *The Lancet* - le journal identifié comme le plus performant en termes de nombre moyen de citations obtenu par article.

ductions. Chaque variable est alors dédoublée en une variable endogène - relative au sous-champ des coronavirus - et une variable exogène - relative au reste du champ scientifique biomédical.

Pour caractériser la réussite présente d'un auteur, nous utilisons son nombre de contributions au cours de l'année considérée et aussi le nombre de citations que les contributions de cette année représentent. Nous expliquons cette réussite présente par une accumulation de capitaux dans le passé : les variables explicatives, afin de représenter une trajectoire, sont en général la somme des métriques de productions et de reconnaissance des travaux de l'auteur sur ses années d'activité antérieures. Nous recourons au nombre total d'articles publiés et de citations reçues sur les travaux antérieurs à l'année considérée dans ce cadre. De même, le nombre total de co-auteurs utilisé dans notre modélisation empirique est constitué du nombre d'auteurs - uniques - avec qui l'auteur a collaboré antérieurement à l'année évaluée. Nous construisons de plus un indice du prestige de publication pour chaque auteur : celui-ci dépend du nombre d'articles publiés par l'auteur et du journal de publication. Nous calculons pour chaque journal la moyenne C_j des citations obtenues par les articles du journal j sur notre période d'étude. Pour chaque auteur i , le score de prestige de publication correspond à $\sum_{j \in J} A_{ij} \times C_j$ où A_{ij} désigne le nombre d'articles écrits par l'auteur i dans le journal j . Etant donné que l'on différencie les articles entre appartenant au champ des coronavirus ou non, chaque auteur reçoit deux variables pour les métriques détaillées, l'une étant la variable contribué au capital endogène, l'autre au capital exogène.

5.2 Approches empiriques

Pour chaque année t , nous établissons la relation linéaire suivante, déterminée avec une régression MCO :

$$Y_{it} = \gamma_t + \sum_{j \in J} (\alpha_{tj} KX_{itj} + \beta_{tj} KE_{itj}) + \epsilon_{it} \quad (1)$$

où KE_{it} et KX_{it} désigne les variables de production scientifiques de l'individu i calculées sur la période $[t-h, t-1]$, où $t-h$ désigne la première année d'apparition de l'individu i dans notre base (la date du premier article scientifique jamais publié par l'auteur i sur Scopus). Les KX_{ij} sont les variables mesurant le capital exogène (variables mesurant la production scientifiques hors du sous-champ de la recherche sur les coronavirus) et les KE_{ij} le capital endogène (variables mesurant la production scientifique au sein du sous-champ de la recherche sur les coronavirus). Y_{it} désigne le succès de l'individu i lors de l'année t au sein du sous-champ de

la recherche sur les coronavirus. Nous estimons les coefficients $(\alpha_j)_{j \in J}$ et $(\beta_j)_{j \in J}$ pour chaque année t .

Afin de comparer les coefficients de nos régressions entre eux et d'une année à l'autre, nous standardisons⁶ nos variables (dépendante et indépendantes). Nous estimons alors les coefficients de régressions standardisés selon le modèle (1). Cette procédure nous permet de changer l'unité de mesure propre à chacune des variables de nos régressions (effectuées pour une année t) en unités de déviation standard, ce qui rend comparable nos coefficients. Les coefficients estimés sont à interpréter comme des effets marginaux : la variation d'une unité standard pour une variable indépendante exogène j - les autres variables indépendantes restant constantes - induit une variation de β_j déviation standard de notre variable dépendante.

5.3 Robustesse

Pour pouvoir s'assurer de la robustesse de nos résultats, nous testons différentes spécifications à la régression (1). Tout d'abord, chacune des variables dépendantes et indépendantes ont été spécifiées tel quel ainsi qu'au logarithme. Aussi, nous mesurons $Y_{i,t}$ - le succès de l'auteur i l'année t - de différentes façons : le nombre d'articles J publiés par l'auteur i l'année t ainsi que le nombre de citations obtenus par ces articles publiés en t (les C_{ijt}). Finalement, nous retenons une variable synthétique qui sera par la suite notre variable d'étude principale où $Y_{it} = \sum_{j \in J} \log(C_{ijt} + 1)$. Ainsi, Y_{it} est positivement corrélé à une hausse du nombre de publications et du nombre de citations.

De même, nous faisons plusieurs tests en restreignant nos observations. Notamment, nous testons nos régressions en restreignant nos observations (i, t) en fonction de leur ancienneté dans le champ des coronavirus (temps écoulé depuis la première publication dans le thème des coronavirus). Ainsi, nous sommes capables de se détacher de la modélisation en milieu continu et de se ramener aux idéaux-types *insider* - *outsider*.

6 Résultats

6.1 Considérations globales

Le résultat le plus saillant de nos régressions est l'absence de significativité des variables exogènes (nombre de publications hors du champ des coronavirus, nombre de citations obtenus par ces mêmes articles, etc.) sur la réussite au sein du champ des coronavirus (cf. Figure 1). Au contraire, la plupart de nos variables endogènes ont un impact sur la réussite, en tout temps de notre période d'étude. Si le nombre d'ar-

6. Nous recourons à la standardisation classique, soustrayant à nos valeurs la moyenne de la variable et les divisant par la déviation standard.

articles publiés au cours d’une carrière ainsi que le prestige des journaux dans lesquels ceux-ci sont publiés - *cf.* le prestige des journaux est mesuré par la moyenne du nombre de citations obtenu par ceux-ci au cours de notre période d’étude - ont un impact positif sur la réussite, on ne peut faire une conclusion similaire pour les autres variables d’étude.

En effet, l’ancienneté a un impact négatif sur la réussite au sein du champ des coronavirus. Ce résultat est contre-intuitif, car l’on pourrait s’attendre à un impact positif de l’expérience. Cette relation peut se comprendre par une prime à la contemporanéité : sur toutes les années d’études, l’ancienneté de participation au champ des coronavirus a un effet négatif sur la réussite présente - ou tout au mieux non significativement différent de 0. D’ailleurs, l’ancienneté dans le champ scientifique hors-coronavirus ne semble pas avoir un impact significatif sur la réussite future dans le champ des coronavirus. Nos tests de robustesse ne trouvent pas de relation avec la variable ancienneté mise au carré. On peut d’ailleurs questionner la pertinence de la théorisation d’un *champ* des coronavirus, dans la mesure où le champ tel que défini empiriquement par notre travail capte un très fort renouvellement annuel. La plupart des auteurs publiant dans le champ des coronavirus sont en effet des primo-auteurs (*cf.* Figure 3). Ainsi, conceptualiser un espace structuré où des auteurs maintiendraient des positions de domination dans le champ ne semble pas totalement en accord avec nos observations - ou du moins, ce champ ne serait pas structuré selon un critère d’expérience.

Le nombre de citations jamais obtenues par les articles d’un auteur a un effet plus ambigu sur la réussite. En effet, nous trouvons une relation quadratique (que nous contrôlons par l’introduction de citations au carré dans les régressions). Avant de se lancer dans l’interprétation, il faut remarquer que le nombre de citations a exactement le même impact sur la réussite d’un chercheur que le nombre d’auteurs avec lesquels il aurait co-écrit, et réciproquement pour les variables citations et coauteurs mises au carré (*cf.* Figure 9). En effet, les articles les plus importants sont sûrement ceux qui ont connu le plus grand nombre de citations. Les relations quadratiques indiquent alors un effet positif de ces deux variables sur la réussite si elles ont des valeurs faibles, un effet négatif pour des valeurs moyennes, et à nouveau positif pour des valeurs élevées. Autrement dit, se faire citer a un rendement décroissant sur la réussite future, à l’exception d’un seuil *superstar* où le chercheur possède un nombre anormalement élevé de citations.

Il est intéressant de noter par ailleurs que publier des articles hors du champ des coronavirus a un impact négatif sur la réussite future. En effet, s’investir fortement hors du champ des coronavirus peut signifier un désintérêt pour ce sous-champ de la part du chercheur,

et diminuer la probabilité que celui-ci se réinvestisse un jour dans un projet de recherche lié au thème des coronavirus.

6.2 Confrontation des résultats empiriques au modèle théorique

Nous pouvons conclure que l’hypothèse **H1** est valide dans la mesure où le nombre d’articles publiés par un auteur dans le champ des coronavirus, le nombre de citations obtenus par ces articles, et le prestige de leur journal de publication, ont un impact positif la réussite de celui-ci au sein du champ des coronavirus. Un *outsider* ne possède donc pas de ces avantages, au contraire d’un *insider*. Au contraire, comme les variables exogènes (relatives à la publication d’article hors du champ des coronavirus) n’ont pas ou peu d’influence sur la réussite au sein du champ des coronavirus, **H2** n’est pas vérifiée. L’intégration différenciée au sous-champ des coronavirus n’est que peut expliquée par des éléments du champ scientifique biomédical hors coronavirus. On peut supposer que d’autres mécanismes sont à l’œuvre, qui dépassent les seules caractéristiques des productions scientifiques : Squazzoni *et al.* (2021) ont pu montrer que durant la première vague pandémique de 2020, le genre est structurant dans la participation au champ scientifique, au détriment des femmes ; ce biais de genre ne se limite pas aux périodes de demande intense de savoir mais est généralisée (Larivière *et al.* (2013)).

L’étude dynamique des relations entre nos variables explicatives et la réussite au sein du champ de la recherche des coronavirus montre une évolution des mécanismes étudiés. Notamment, les chocs de demande de savoir, en particulier ceux correspondant à l’épidémie du SARS et du COVID-19, ont tendance à modifier l’ensemble des relations entre variables au sein de nos régressions. En particulier, il augmente la tendance à la réussite des auteurs possédant un nombre de citations élevé à leur actif au sein du champ des coronavirus. Pour le cas spécifique du COVID-19, nous interprétons la baisse de l’importance du nombre d’articles à l’actif des chercheurs dans l’importance de la réussite par l’explosion du nombre d’articles dans le champ : face à une littérature inexplorable dans son entièreté dans des délais raisonnables, le nombre de citations obtenus par les articles est un signal fort pour un chercheur en quête d’information. Au contraire, dans un monde où la recherche s’est mis à fonctionner à plein régime avec des temps de publications grandement diminués, les journaux et le prestige qui leur sont associés perdent de leur importance pour structurer le champ. Dans la mesure où, de surcroît, les variables exogènes n’ont qu’un très faible ou nul impact sur la réussite future, nous pouvons non seulement réfuter **H3** mais aussi souligner que les relations postulées sont inverses à celles obser-

vées empiriquement. L'entrée massive d'*outsiders* dans le champ a tendance à renforcer - ou plutôt à instituer

- une position dominante des *insiders* les plus dotés en capitaux.

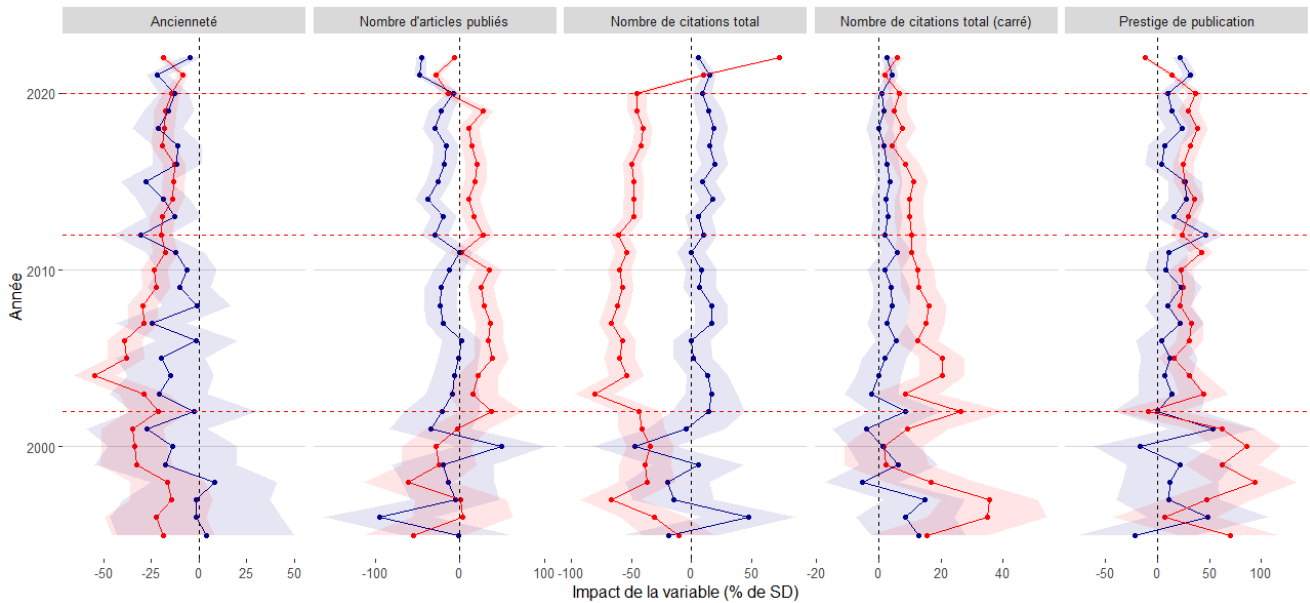


FIGURE 1 – Effets des capitaux endogène (rouge) et exogène (bleu) par année sur le nombre de citations (somme des logarithmes du nombre de citations obtenu par chaque article) d'un auteur.

Note : Chaque point est le coefficient de la variable indiquée en colonne tel qu'obtenu dans la régression (1). Les variables de la régression ont été standardisées et les effets se mesurent en pourcentage de déviation standard (SD). Les traits rouges en pointillés indiquent les épidémies du SARS (2002), du MERS (2012) et du COVID (2020).

Lecture : En 2002, au moment de l'épidémie du SARS, une variation d'une déviation standard de l'ancienneté d'un auteur dans le champ des coronavirus fait baisser de 25%, en moyenne et toutes choses égales par ailleurs, la déviation standard du nombre de citations obtenues par les articles d'un auteur publiés cette année-là (mesurées par la somme des log du nombre de citations obtenus pour chaque article publié).

7 Discussion

Nos analyses sont d'abord confrontées à des limites générées par la base de données Scopus elle-même. Nous avons par exemple personnellement constaté que les identifiants d'auteur étaient parfois mal renseignés : un même auteur peut avoir plusieurs identifiants Scopus différents et même assurément lorsqu'un.e individu.e prend le nom de son.sa conjoint.e. Il.elle est alors dédoublé.e et son parcours professionnel tronqué dans notre base d'étude (et créer un biais lié au genre de l'auteur.trice). Une extension de notre travail peut consister à répliquer les analyses que nous avons conduites sur d'autres base de données scientométriques afin de s'assurer de la robustesse de nos résultats.

Il est par ailleurs envisageable que nos choix de sélection et retraitement des données introduisent des biais. Le choix de se limiter à certains journaux, bien qu'important pour avoir un champ d'étude homogène et comparable en termes de contributions, pourrait masquer un grand nombre de phénomènes caractéristiques des publications scientifiques faites dans des revues moins prisées.

Sur le plan théorique, notre étude repose sur des hypothèses fortes. En premier lieu, nous supposons l'existence d'un champ scientifique, ce qui n'est pas une évi-

dence lorsque l'on considère l'hétérogénéité du monde de la recherche scientifique, et l'hétérogénéité des intensités de connexions entre groupes de chercheurs. En effet, d'autres fragmentations que le thème de recherche peuvent rentrer en jeu, tels que la situation géographique, la langue, le laboratoire, les réseaux sociaux, l'éducation, *etc.* En second lieu, réciproquement, la théorisation d'un sous-champ de la recherche sur les coronavirus n'a pas été vérifiée empiriquement. Bien que cette hypothèse soit loin d'être absurde, elle repose sur l'idée que les coronavirus sont un objet d'étude distinct de tous les autres - des autres virus par exemple - et que tout capital scientifique exogène, accumulé en-dehors de ce sous-champ, ne serait pas lui-même composé d'une certaine part de capital endogène. En particulier, la Figure 2 montre une évolution concomitante du nombre d'articles publiés par les auteurs de notre échantillon au sein du champ de la recherche des coronavirus et en-dehors de celui-ci. On peut expliquer le phénomène ou bien par la crise du COVID qui aurait eu un impact sur l'ensemble du monde scientifique, tout champ, ou bien nous captions en-dehors du champ des coronavirus des articles apparentés coronavirus et qui par conséquent bénéficient aussi du choc de demande de savoir.

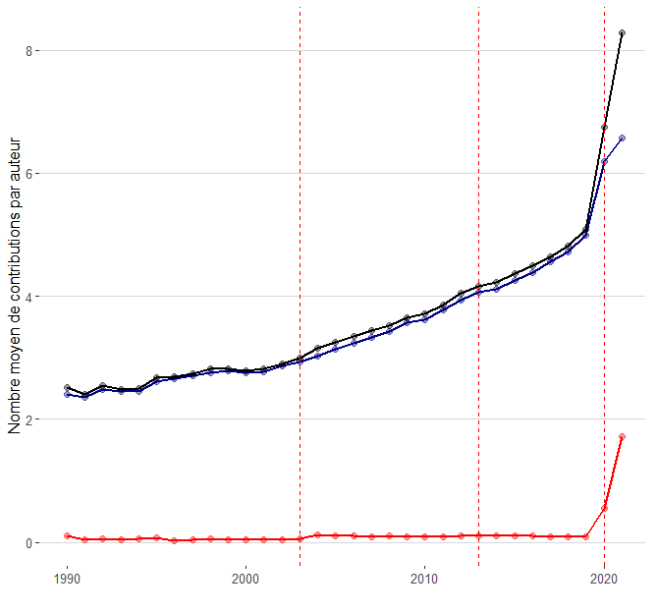


FIGURE 2 – Nombre moyen de contributions par auteur et par année dans notre échantillon

Note de lecture : le nombre moyen de contributions par auteurs en 2021 est de 8,3 (ligne noire). En moyenne, 1,7 contributions relèvent des coronavirus (ligne rouge) tandis que 6,3 n'en relèvent pas (ligne bleue).

Bien que difficilement opérationnalisable, il serait alors envisageable de caractériser les autres sous-champs scientifiques par une proximité plus ou moins importante au sous-champ des coronavirus. La plus ou moins grande transférabilité du capital exogène d'un auteur au champ des coronavirus pourrait être interprétée à cette aune, *i.e.* tous les capitaux exogènes que nous caractérisons ici comme équivalents ne se valent probablement pas pour réussir au sein du sous-champ des coronavirus.

Nous avons susmentionné notre absence de données quant aux dates de citation de chaque article (nous n'avons à notre disposition que le nombre total de citations de l'article x lors de l'extraction de nos données en mars 2022, et nous ne sommes pas en mesure de savoir par qui, quand et comment cet article a été cité). Dans l'hypothèse où chaque article d'une année t est impacté de la même façon par ce biais, cela ne pose pas de problème dans l'interprétation des analyses causales. Cependant, l'éventualité existe aussi que le succès d'un auteur à partir d'une année t augmente le succès de ses publications publiées lors des années $t - h$. Ce potentiel biais d'endogénéité - causalité inverse - ne concerne cependant que le nombre de citations de la production d'un auteur, et non son volume de publication.

Finalement, nous manquons d'un certain nombre de variables de contrôle qui nous auraient permis de

conclure de façon tranchée sur la véracité de nos hypothèses de travail. Tout d'abord, au sein de nos régressions, nous ne contrôlons pas un nombre conséquent de variables biographiques des auteurs. En particulier, l'éducation / le diplôme, le genre, le laboratoire d'appartenance, le rang au sein de la hiérarchie de ce laboratoire, le talent / l'efficacité productive intrinsèque de l'individu, ne sont pas observés. De même, si nous avons exploité le caractère ponctuel des chocs de demande de savoir, ce qui permet de comparer une année consécutive à l'autre par expérience naturelle, nous ne sommes pas à l'abri d'autres évolutions impactant simultanément la structure du champ. Typiquement l'épidémie de COVID-19 a affecté le monde de la recherche scientifique à travers un choc de demande certes, mais a aussi eu des énormes conséquences sur le mode de fonctionnement de nos sociétés, en banalisant le télétravail notamment, et par conséquent les interactions par télécommunication, en bouleversant le rythme de travail et de publication des chercheurs, à cause des confinements et des redistributions des moyens de la médecine vers les hôpitaux. Nous aurions alors pu exploiter le caractère ciblé du choc de demande sur les coronavirus en définissant un autre sous-champ scientifique biomédical disjoint que l'on aurait utilisé pour contrôler les évolutions structurelles du sous-champ de la recherche sur les coronavirus.

8 Conclusion

La modélisation proposée dans cet article contribue à la compréhension de la restructuration du champ scientifique à la suite d'un choc de demande de savoir sur les dynamiques de réussite au sein d'un champ d'étude scientifique. Les propriétés du champ de recherche sur les coronavirus, champ de recherche confronté à de conséquentes évolutions au cours des 20 dernières années, nous ont permis d'identifier des mécanismes internes quant à l'organisation de champ. En particulier, lorsque la demande de savoir s'intensifie, on observe une hausse importante du nombre de primo-auteurs dans le champ, qui doivent se confronter à l'ascendance hiérarchique scientifique des auteurs déjà installés. C'est particulièrement le cas à la suite de la crise du COVID-19 qui semble restructurer le champ fortement autour des auteurs insérés dans le domaine peu avant l'explosion pandémique. Nos résultats, s'ils permettent de conclure au sein du champ d'étude, justifient une méthode plus robuste, faisant notamment intervenir un autre champ de recherche scientifique comme groupe de contrôle.

Références

- AVIV-REUVEN, S. et ROSENFELD, A. (2021). Publication patterns' changes due to the COVID-19 pandemic : a longitudinal and short-term scientometric analysis. *Scientometrics*, 126(8):6761–6784.
- BAAS, J., SCHOTTEN, M., PLUME, A., CÔTÉ, G. et KARIMI, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1):377–386.
- BOURDIEU, P. (1976). Le champ scientifique. *Actes de la recherche en sciences sociales*, 2(2):88–104.
- FORTUNATO, S., BERGSTROM, C. T., BÖRNER, K., EVANS, J. A., HELBING, D., MILOJEVIĆ, S., PETERSEN, A. M., RADICCHI, F., SINATRA, R., UZZI, B., VESPIGNANI, A., WALTMAN, L., WANG, D. et BARABÁSI, A.-L. (2018). Science of science. *Science*, 359(6379):eaao0185.
- JÆGER, M. M. et BREEN, R. (2016). A Dynamic Model of Cultural Reproduction. *American Journal of Sociology*, 121(4):1079–1115.
- LARIVIÈRE, V., GINGRAS, Y., SUGIMOTO, C. R. et TSOU, A. (2015). Team size matters : Collaboration and scientific impact since 1900 : On the Relationship Between Collaboration and Scientific Impact Since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332.
- LARIVIÈRE, V., NI, C., GINGRAS, Y., CRONIN, B. et SUGIMOTO, C. R. (2013). Bibliometrics : Global gender disparities in science. *Nature*, 504(7479):211–213.
- LATOUR, B., WOOLGAR, S. et LATOUR, B. (2013). *La vie de laboratoire : la production des faits scientifiques*. Numéro 18 de La Découverte/Poche Sciences humaines et sociales. La Découverte, Paris, nachdr. édition.
- LU WANG, L., LO, K., CHANDRASEKHAR, Y., REAS, R., YANG, J., EIDE, D., FUNK, K., KINNEY, R., LIU, Z., MERRILL, W., MOONEY, P., MURDICK, D., RISHI, D., SHEEHAN, J., SHEN, Z., STILSON, B., WADE, A. D., WANG, K., WILHELM, C., XIE, B., RAYMOND, D., WELD, D. S., ETZIONI, O. et KOHLMEIER, S. (2020). CORD-19 : The Covid-19 Open Research Dataset. *ArXiv*, page arXiv :2004.10706v2.
- PETERSEN, A. M., FORTUNATO, S., PAN, R. K., KASKI, K., PENNER, O., RUNGI, A., RICCABONI, M., STANLEY, H. E. et PAMMOLLI, F. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences*, 111(43):15316–15321.
- SQUAZZONI, F., BRAVO, G., GRIMALDO, F., GARCÍA-COSTA, D., FARJAM, M. et MEHMANI, B. (2021). Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic. A study on 2329 Elsevier journals. *PLOS ONE*, 16(10):e0257919.

9 Annexes

Les données et *scripts* de code pour répliquer les analyses sont disponibles sur le répertoire *GitHub* suivant : <https://github.com/mthsansu/CovidLiterature>.

TABLE 1 – Liste des journaux sélectionnés et période(s) associée(s)

Journal	Période(s)
<i>Journal of Virology</i>	1 ; 2 ; 3
<i>Advances in Experimental Medicine and Biology</i>	1
<i>Virology</i>	1 ; 2 ; 3
<i>Journal of General Virology</i>	1 ; 2 ; 3
<i>Archives of Virology</i>	1 ; 3
<i>Plos One</i>	1 ; 2 ; 3 ; 4
<i>Avian Diseases</i>	1
<i>American Journal of Veterinary Research</i>	1
<i>Veterinary Microbiology</i>	1 ; 3
<i>Virus Research</i>	1 ; 2 ; 3
<i>Veterinary Record</i>	1
<i>Journal of Immunology</i>	1 ; 2
<i>Veterinary Immunology and Immunopathology</i>	1
<i>Journal of Virological Methods</i>	1 ; 2 ; 3
<i>Journal of Biological Chemistry</i>	1 ; 2
<i>Emerging Infectious Diseases</i>	2 ; 3
<i>Vaccine</i>	2 ; 3
<i>Journal of Clinical Microbiology</i>	2
<i>Lancet</i>	2
<i>Proceedings of the National Academy of Sciences of the United States of America</i>	2
<i>Biochemical and Biophysical Research Communications</i>	2
<i>Journal of Infectious Diseases</i>	2
<i>Scientific Reports</i>	3 ; 4
<i>Viruses</i>	3 ; 4
<i>Plos Pathogens</i>	3
<i>Antiviral Research</i>	3
<i>Virology Journal</i>	3
<i>International Journal of Environmental Research and Public Health</i>	4
<i>Sustainability Switzerland</i>	4
<i>Frontiers in Psychology</i>	4
<i>BMJ</i>	4
<i>Frontiers in Immunology</i>	4
<i>Frontiers in Public Health</i>	4
<i>International Journal of Molecular Sciences</i>	4
<i>Journal of Medical Virology</i>	4
<i>ACM International Conference Proceeding Series</i>	4
<i>BMJ Open</i>	4
<i>Journal of Clinical Medicine</i>	4
<i>Vaccines</i>	4

Note de lecture : un journal est sélectionné s'il est parmi les 15 qui contribuent le plus dans le domaine des coronavirus en termes de nombres d'articles au sein d'une période donnée. Un journal peut être parmi les 15 plus importants sur plusieurs périodes. Par exemple, le *Journal of Virology* est dans les 15 plus gros contributeurs sur les coronavirus durant les périodes 1, 2 et 3, mais pas durant la période 4.

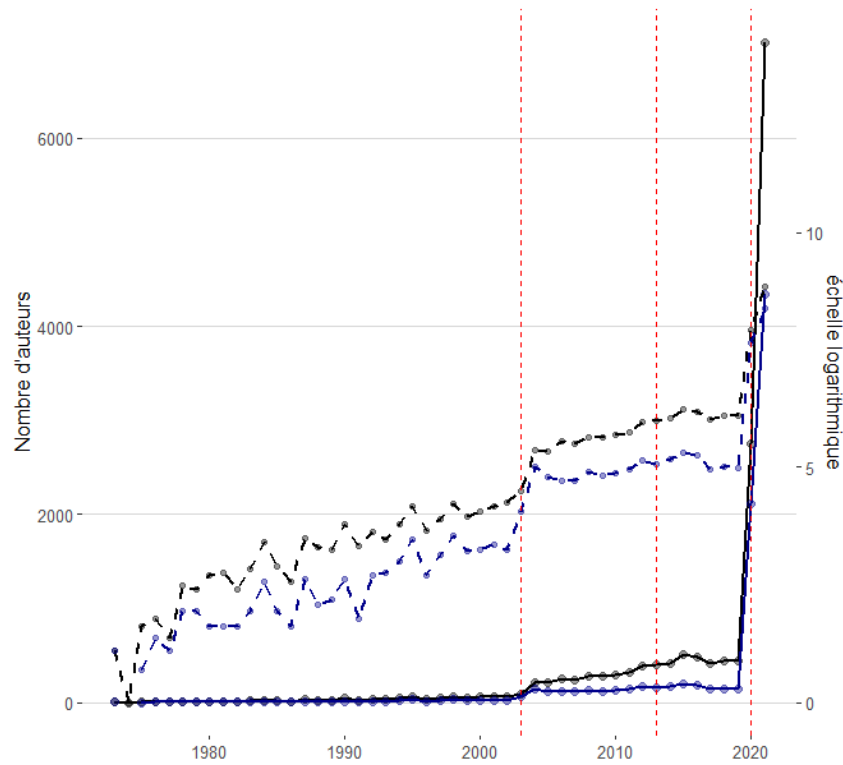


FIGURE 3 – Nombres d’auteurs et de primo-auteurs par année contribuant au sous-champ de recherche des coronavirus dans notre échantillon.

Note de lecture : la ligne pleine noire donne le nombre absolu d’auteurs contribuant par année au sous-champ. La ligne pleine bleue correspond au nombre de primo-auteurs dans le sous-champ, c’est-à-dire aux *outsiders*. Les lignes pointillées correspondent aux valeurs logarithmiques et se réfèrent à l’axe des ordonnées à droite. La crise du COVID-19 a bouleversé complètement le sous-champ, là où la crise du SARS semblait auparavant l’avoir ouvert. La crise du MERS ne semble pas avoir eu de réponse scientifique franche dans notre échantillon.



FIGURE 4 – Part des primo-auteurs dans le sous-champ des coronavirus par année dans notre échantillon.

Note de lecture : en 2020, la part d’*outsiders* parmi les auteurs du domaine des coronavirus s’élève à 0,77. On observe que le SARS et le COVID-19 semblent avoir occasionné une rentrée importante de primo-auteurs dans le sous-champ.

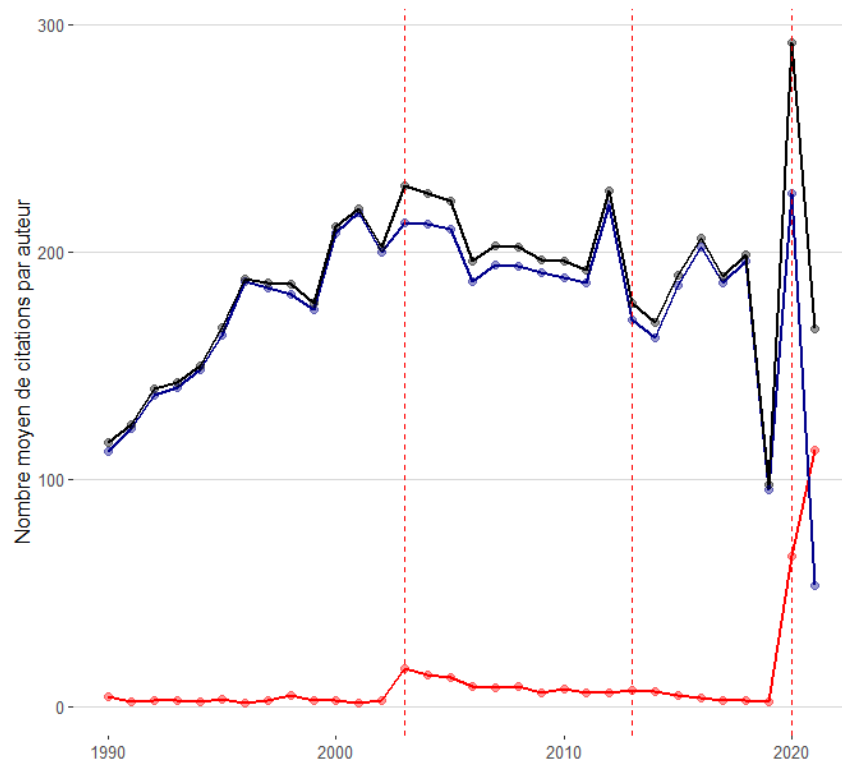


FIGURE 5 – Nombre moyen de citations par auteur pour ses articles publiés à une année donnée dans notre échantillon.

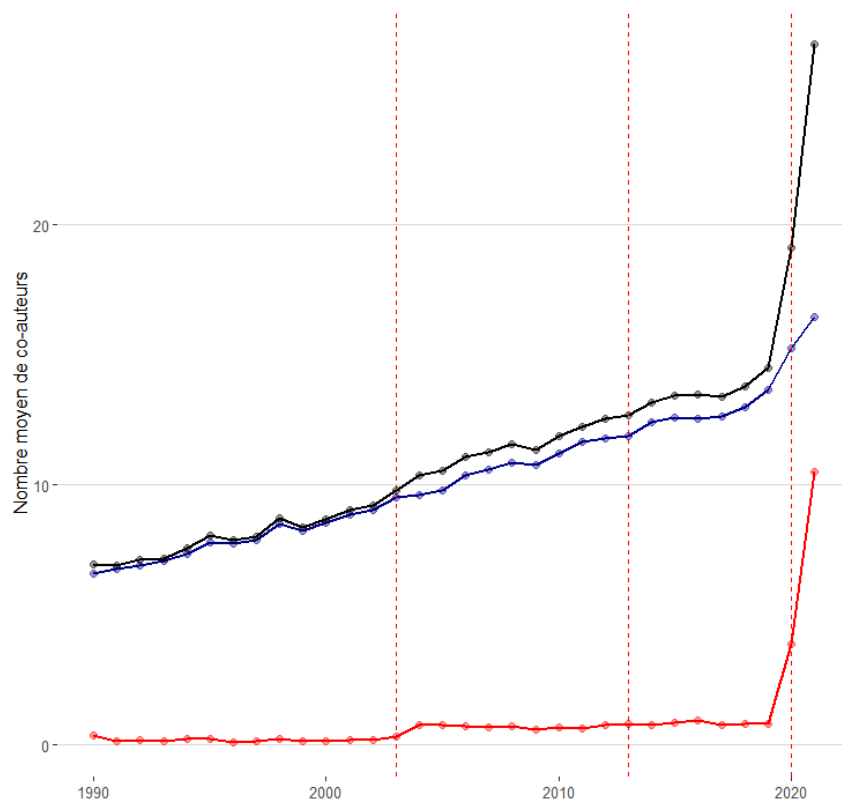


FIGURE 6 – Nombre moyen de citations par auteur pour ses articles publiés à une année donnée dans notre échantillon.

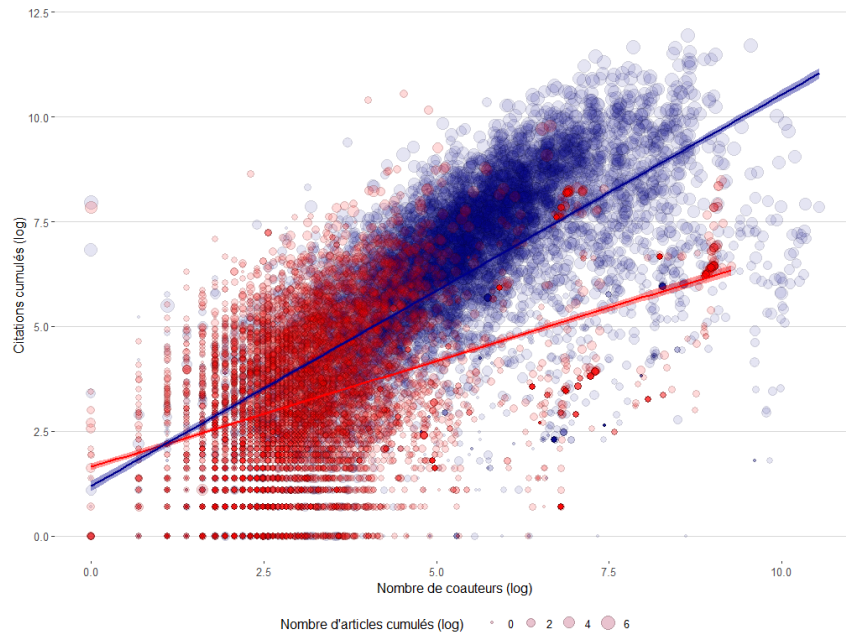


FIGURE 7 – Relation entre le nombre de co-auteurs (log) et le nombre de citations cumulées (log) par auteur.

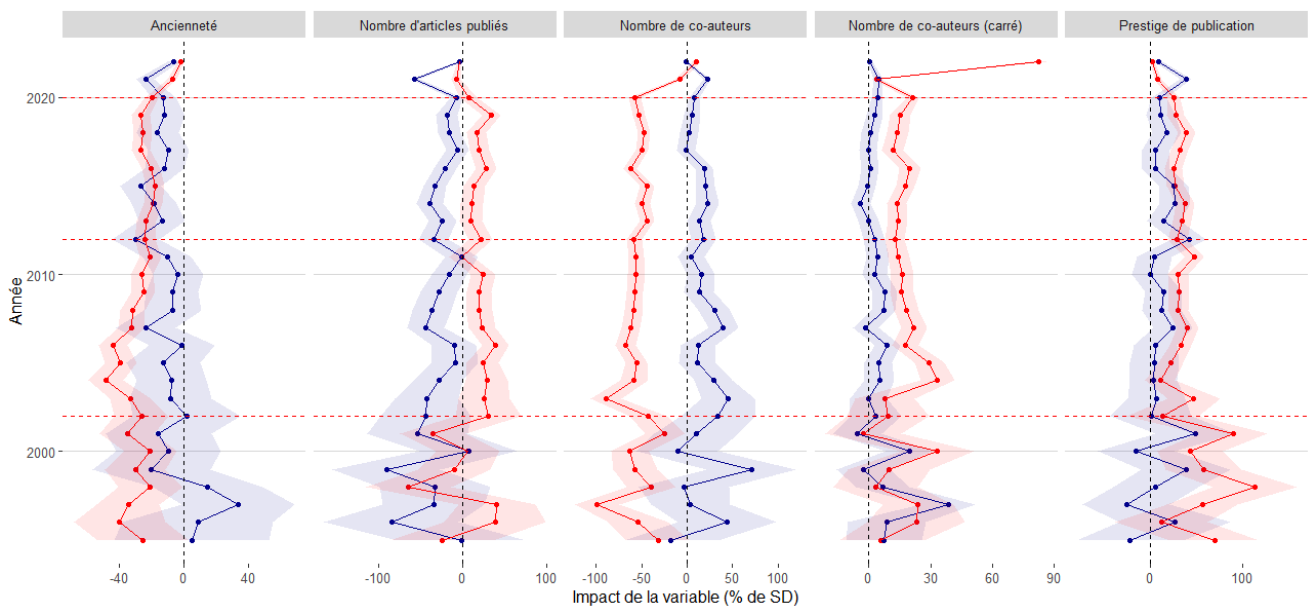


FIGURE 8 – Effets des capitaux endogène (rouge) et exogène (bleu) par année sur le nombre de citations (somme des log du nombre de citations obtenu par chaque article) d'un auteur.

Note : le modèle est similaire au modèle (1) dont les résultats sont retranscrits dans la [Figure 1](#) à la différence du nombre total de citations qui est remplacé par le nombre total de co-auteurs. Les motifs liés au modèle (1) sont robustes.

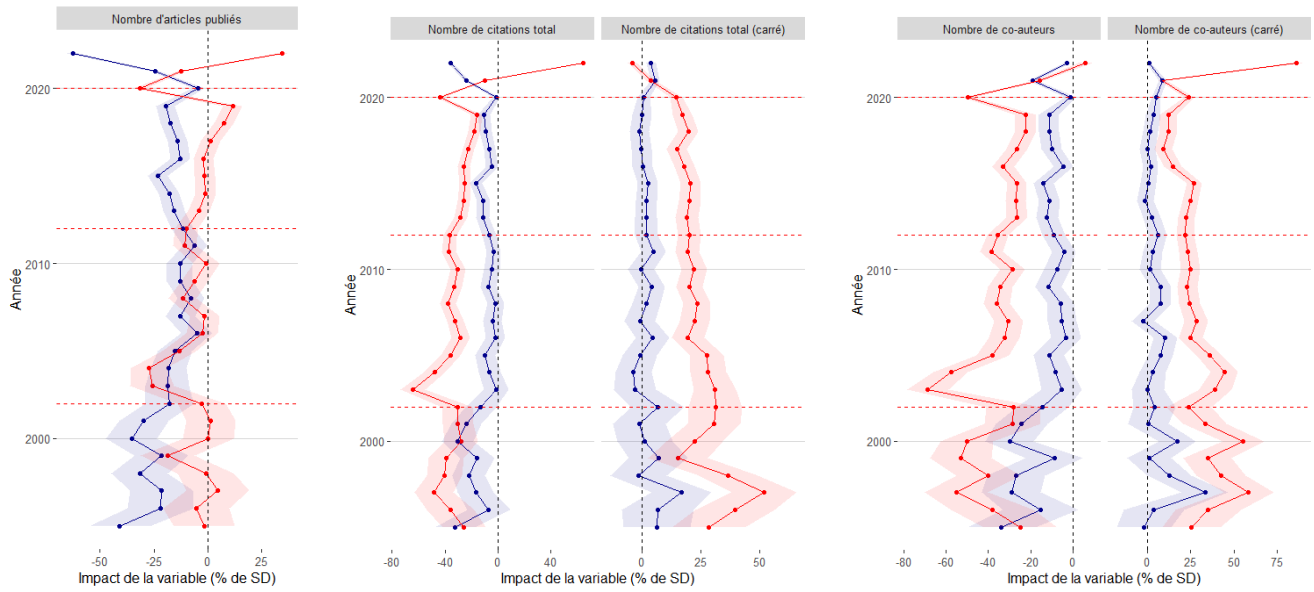


FIGURE 9 – Tests de robustesse :

- (a) : Effets du nombre d'articles dans le champ des coronavirus (rouge) et hors de ce champ (bleu) par année sur le nombre de citations (somme des log du nombre de citations obtenu par chaque article) d'un auteur.
- (b) : Effets du nombre total de citations dans le champ des coronavirus (rouge) et hors de ce champ (bleu) par année sur le nombre de citations (somme des log du nombre de citations obtenu par chaque article) d'un auteur.
- (c) : Effets du nombre total de citations dans le champ des coronavirus (rouge) et hors de ce champ (bleu) par année sur le nombre de citations (somme des log du nombre de citations obtenu par chaque article) d'un auteur.

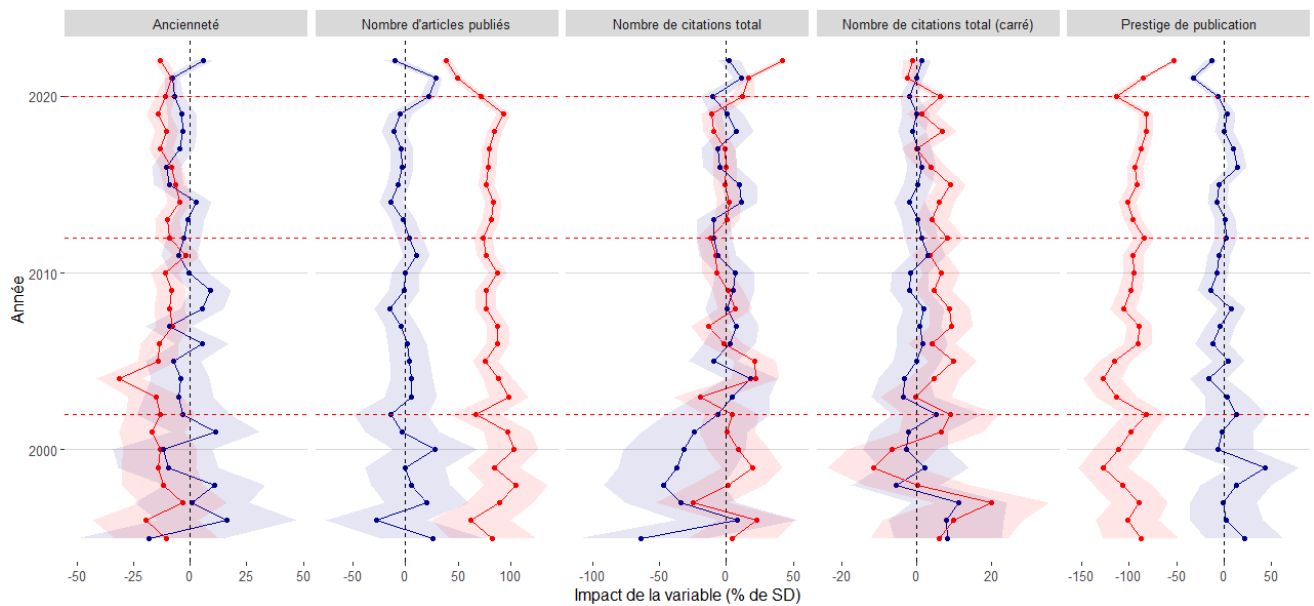


FIGURE 10 – Effets des capitaux endogène (rouge) et exogène (bleu) par année sur le nombre de publications d'un auteur.

Note : la variable dépendante du modèle est le nombre de publications dans le domaine des coronavirus d'un auteur sur l'année considérée. A l'exception de la variable de prestige de publication, les motifs liés au modèle initial sont similaires.

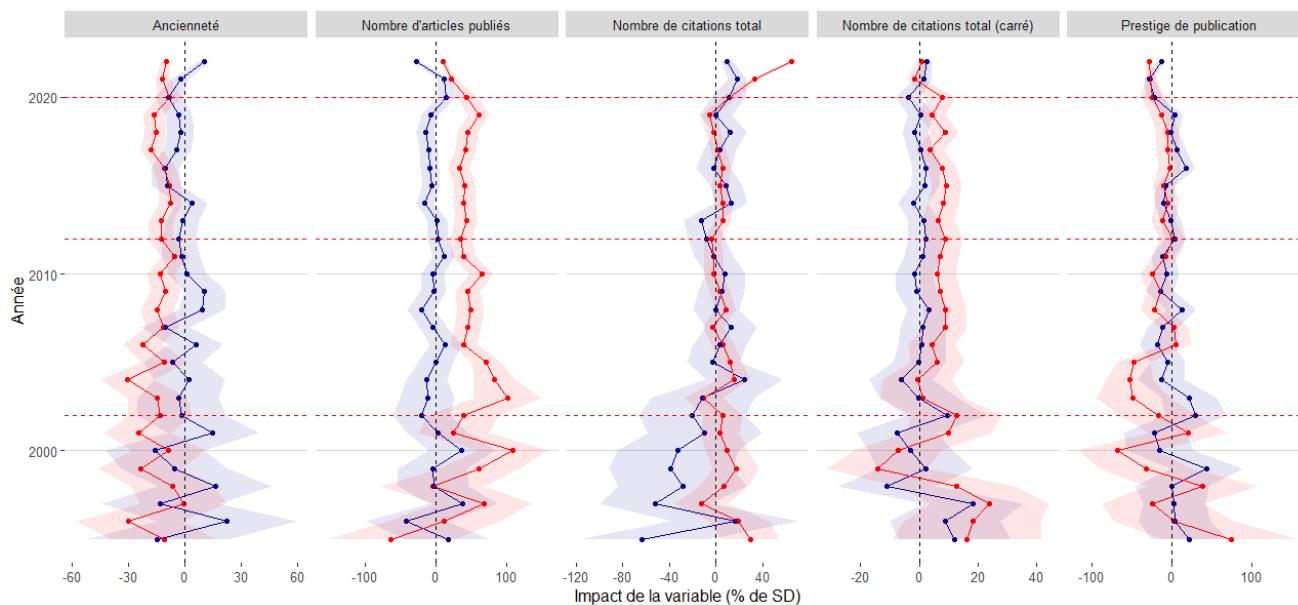


FIGURE 11 – Effets des capitaux endogène (rouge) et exogène (bleu) par année sur le nombre de citations (somme des log du nombre de citations obtenu par chaque article) d'un auteur.

Note : le modèle est similaire au modèle (1) dont les résultats sont retranscrits dans la [Figure 1](#) à la différence que nous restreignons notre échantillon d'observations aux seuls auteurs ayant déjà du capital endogène sur l'année considérée.