

Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique

Pierre-Henri Bono
Sandra Debu
Rozenn Desplatz
Maxime Hayet
Clément Lacouette-Fougère
Alain Trannoy



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

Sommaire

Résumé.....	2
Introduction.....	3
1. La base de données bibliométriques	5
1.1. Présentation de la base EconLit.....	5
1.2. Sélection des articles d'évaluation d'impact	6
1.3. Création de la base de travail	9
2. Les résultats.....	11
2.1. Près de 7 500 articles d'évaluation d'impact pour tous les pays	11
2.1.1. Une production mondiale dominée par les États-Unis	12
2.1.2. Des évaluations portées par l'économie du développement et centrées sur les secteurs de l'emploi, de la santé et de l'éducation.....	16
2.1.3. Des articles d'évaluation publiés dans des revues généralistes en économie ..	17
2.2. Par rapport aux pays les plus avancés, une production française relativement modeste, mais en rattrapage	18
2.2.1. Une dynamique positive dans tous les pays de l'échantillon, en particulier en France.....	19
2.2.2. Des thématiques de recherche communes : emploi, santé, éducation	21
2.2.3. Des évaluations publiées majoritairement dans les revues nationales (hors monde anglo-saxon).....	24
2.2.4. Les grands acteurs académiques nationaux très présents sur le champ de l'évaluation	25
Références	33
Annexes.....	35
A. Codes JEL de catégorie 1	35
B. Classement des journaux.....	36
C. Graphiques avec les États-Unis	36
D. Évolution rapportée à la population	37
E. Top 20 des institutions par pays	38
F. Tests sur la base d'articles relative à la France	45

Résumé

À l'aide de la base EconLit, nous réalisons une étude bibliométrique pour quantifier la production d'évaluations d'impact de politiques publiques en France et dans six autres pays. Les évaluations d'impact se sont largement développées dans tous les pays étudiés à partir des années 2000 (et bien avant pour les États-Unis). La France présente une production relativement modeste, mais connaît une accélération dans la période récente, ce qui lui permet de rattraper son retard vis-à-vis d'autres pays plus avancés comme l'Allemagne ou le Royaume-Uni. Les évaluations dans les pays européens couvrent majoritairement les domaines de l'emploi, de l'éducation et de la santé. Ces évaluations sont dans la plupart des cas produites par les principaux acteurs universitaires et de recherche nationaux ; la France se démarque cependant par la présence importante d'acteurs à mi-chemin entre le monde de la recherche et l'administration.

Mots clés : Comparaison internationale, étude d'impact, évaluation, politique publique.
Codes JEL : A1 ; D04 ; H00 ; C2.

Vingt ans d'évaluations d'impact en France et à l'étranger

Analyse quantitative de la production scientifique

Pierre-Henri Bono
Sciences Po / CEVIPOF

Sandra Debu
France Stratégie

Rozenn Desplatz
France Stratégie

Maxime Hayet
Direction interministérielle de la transformation publique

Clément Lacouette-Fougère
Direction interministérielle de la transformation publique

Alain Trannoy
EHESS et France Stratégie

Introduction

Évaluer les conséquences des politiques publiques et connaître leur impact sur les bénéficiaires est un enjeu démocratique majeur. Dans un contexte marqué par un niveau élevé de dépenses publiques conjugué à une défiance croissante des citoyens envers les institutions politiques, l'évaluation d'impact entend administrer la preuve que les politiques publiques ont atteint leurs objectifs (Desplatz, Ferracci, 2016). Cette information doit aider les décideurs publics à éclairer leurs choix lors de la conception et de la mise en œuvre de nouvelles politiques publiques, à arbitrer entre plusieurs options, mais aussi à étendre ou à supprimer des dispositifs existants en fonction de l'impact démontré sur la trajectoire des agents économiques et des usagers.

Depuis une quinzaine d'années, cette préoccupation de mise à disposition judicieuse des connaissances disponibles pour les décideurs s'est diffusée sous le vocable de politiques publiques fondées sur la preuve, ou « *evidence-based policy* » (Evaluation Gap Working Group, 2006). Issues du domaine de la médecine et des sciences du vivant, ces approches se sont progressivement étendues à l'ensemble des champs d'action publique, en particulier les politiques d'emploi, de lutte contre la pauvreté ou d'éducation (Laurent *et al.*, 2009). Elles ont donné lieu à des débats au sein des pays, tant sur les méthodes que sur leurs finalités (Nutley, Davies, 2001 ; Devaux-Spatarakis, 2015 ; Simha, 2017) et ont suscité l'intérêt des gouvernements et des décideurs ainsi que des mouvements de rapprochement de l'offre et de la demande de preuves (Nutley *et al.*, 2000 ; Sanderson, 2002 ; Evidence Works Conference, 2016).

En France, le développement de l'évaluation d'impact est encore modeste comparé à celui d'autres pays, en particulier anglo-saxons. Même si des progrès ont été réalisés dans le domaine (apparition de nouveaux acteurs, ouverture des données, engagements publics au plus haut niveau par les exécutifs successifs), évaluer les effets d'une politique est loin d'être devenu un réflexe dans notre pays. Trop souvent encore, l'évaluation d'impact est confondue avec d'autres démarches, comme le contrôle, l'audit, le conseil, la consultation citoyenne, etc. Les chercheurs et universitaires investis sur le sujet sont encore peu nombreux et les moyens pour conduire des travaux d'évaluation restent limités, en particulier les financements de la recherche scientifique dédiés au sujet. Dans une note publiée en 2017, l'Institut des politiques publiques (IPP) estimait que seules 13 % des évaluations produites sur la période 2012-2017 en France relevaient de l'évaluation d'impact (Bozio, Romanello, 2017).

Ce bilan, partagé par nombre de défenseurs de l'évaluation, n'offre que peu de réponses à des questions simples que soulève la recherche d'un plus grand déploiement de l'évaluation d'impact en France : combien d'évaluations d'impact sont produites en France ? Comment la France se situe-t-elle par rapport aux pays les plus avancés ? Dans quels secteurs et sur quelles politiques publiques des évaluations sont-elles majoritairement réalisées ? Qui sont les principaux producteurs d'évaluation ?

Notre étude, qui compare la France à six pays (les États-Unis, le Royaume-Uni, la Suède, le Canada, l'Allemagne et les Pays-Bas), entend donc répondre à ces questions. Elle vise plus précisément à comparer la vitalité de la recherche académique en France par rapport à d'autres pays avancés, et à questionner dans un second temps son degré d'utilisation et d'influence sur la décision publique (Bénassy-Quéré, Blanchard, Tirole, 2017). Elle se décline en deux volets complémentaires :

- un premier volet est composé de l'étude bibliométrique présentée ci-après qui analyse la production globale d'évaluations d'impact avant de comparer les sept pays de notre échantillon ;
- un second volet plus qualitatif s'appuie sur la réalisation de fiches monographiques par pays pour examiner la place des évaluations d'impact dans le processus de décision selon trois critères (degré d'ouverture du milieu administratif aux chercheurs ; existence de dispositifs formels favorisant la conduite d'évaluations d'impact ; diffusion et influence des évaluations). Ce second volet donnera lieu à un second rapport dédié.

L'étude bibliométrique permet de quantifier, à partir des articles académiques, la qualité et le volume de la production scientifique. Elle a été mobilisée à plusieurs reprises dans des travaux en économie cherchant à mesurer la production totale d'articles scientifiques (Lubrano *et al.*, 2003), ou à analyser la qualité de la production dans le but d'opérer des classements des centres de recherche (Combes, Linnemer, 2003 ; Bosquet, Combes, 2012), voire d'analyser la productivité de leurs chercheurs (Combes, Linnemer, 2001). À revers de ces travaux, notre analyse s'intéresse au volume d'études produites sur les pays et à leur évolution dans leur temps, sans chercher à évaluer leur qualité ou leur apport scientifique supposé. Elle porte de ce fait sur une période d'analyse longue (une vingtaine d'années), à la différence des travaux précités qui considèrent une période de référence plus courte.

1. La base de données bibliométriques

Dans cette partie, nous présentons la base de données d'articles scientifiques en économie EconLit utilisée pour notre travail. Nous détaillons ensuite comment ont été sélectionnés les articles et enfin quels traitements ont été effectués pour obtenir notre base de travail finale.

1.1. Présentation de la base EconLit

Tout exercice bibliométrique nécessite une base de données recensant la production académique. Parmi les bases de données disponibles¹, notre choix s'est porté sur la base EconLit. Cette base de données rassemble un grand nombre de ressources académiques (articles avec comité de lecture, ouvrages, thèses, documents de travail, etc.) en sciences économiques au sens large (y compris les revues de management et de gestion). Elle a été constituée et développée par l'American Economic Association². Elle est alimentée et mise à jour par les éditeurs de revues qui signalent les nouvelles publications. Les premières entrées de la base datent de 1886 et la base totalise à ce jour près de 1,3 million d'observations. EconLit est également la base de données la plus couramment utilisée dans les travaux bibliométriques que nous avons identifiés (voir ci-dessus).

Pour ce travail, nous limitons notre étude aux seuls articles de recherche afin de disposer d'une base homogène sur la nature et la qualité des travaux produits, reconnue internationalement (Mirucki, 1999). En particulier, nous excluons les documents de travail. La principale raison tient à la qualité scientifique très disparate des documents de travail qui, à la différence des revues à comité de lecture, n'ont pas été soumis à une procédure rigoureuse d'examen par les pairs. Par ailleurs, la base EconLit surreprésente les documents de travail anglo-saxons, ce qui aurait faussé nos analyses en surestimant la production anglo-saxonne au détriment des autres pays. Enfin, la prise en compte des documents de travail aurait créé des doubles comptes, car bon nombre d'entre eux sont ensuite publiés dans des revues académiques. Toutefois, en excluant les documents de travail de notre champ d'analyse, nous ne tenons pas compte de la production académique récente, puisqu'il faut en moyenne environ deux ans³ pour qu'un document de travail soumis à une revue scientifique y soit publié.

D'un point de vue méthodologique, EconLit présente de nombreux avantages, mais aussi quelques limites. La base porte sur la littérature économique et comprend certaines publications en management et gestion. En outre, elle permet de définir des requêtes de recherche complexes sur différents champs d'un article de manière indépendante, comme le titre de l'article, son résumé ou ses mots clés. Cette caractéristique distingue EconLit d'autres bases, en particulier Google Scholar, moins souples dans les possibilités qu'elles offrent de définir les requêtes (Bosquet, Combes, 2012). Par ailleurs, quelle que soit la

¹ Les autres sources de données disponibles que nous avons identifiées sont : RePEc, Web of Science Citation Databases, JSTOR, ScienceDirect, Google Scholar. Parmi ces sources, EconLit est l'unique base de données permettant de faire une recherche sur les seules publications en économie et comprenant un large nombre de revues.

² <https://www.aeaweb.org/econlit/>.

³ Voir Ellison (2002) pour une étude sur la durée de publication.

langue de l'article, le titre et le résumé sont traduits en anglais dans la base⁴, ce qui permet de limiter les requêtes à la seule langue anglaise. Enfin, la base identifie de nombreux champs utiles pour l'analyse comme la langue de l'article, les codes JEL⁵ quasiment systématiquement renseignés⁶, ou le ou les pays sur lesquels porte l'article de recherche. Ce dernier champ revêt une importance considérable dans notre analyse, car il nous permet d'identifier le nombre d'études qui portent sur un pays en particulier.

S'agissant des limites, la base ne recense pas les travaux d'évaluation produits par la sphère non académique (services administratifs ou ministériels, cabinets d'études et de conseil, etc.), fussent-ils scientifiques, c'est-à-dire relevant de la même catégorie de travaux que ceux auxquels on s'intéresse (cf. ci-dessous), dès lors qu'ils ne sont pas publiés dans des revues scientifiques.

1.2. Sélection des articles d'évaluation d'impact

Le but de cette étude est de disposer d'une mesure du nombre d'articles de recherche sur les évaluations d'impact *ex post* au sens large, qui concernent un pays en particulier⁷. Pour cela nous avons défini une requête sur la base EconLit portant sur les titres et les résumés uniquement à partir d'un certain nombre d'expressions identifiant les articles souhaités.

Une des caractéristiques des évaluations d'impact *ex post* est qu'elles sont attachées à une méthode d'identification permettant d'obtenir une mesure causale de l'impact. Si ces méthodes sont plurielles, leur nombre reste suffisamment contenu pour les utiliser dans la requête. Cette dernière apporte également la garantie que les auteurs ont fait usage d'au moins l'une des méthodes robustes de l'évaluation d'impact. Nous avons identifié huit méthodes, plus une catégorie généraliste. Les méthodes sont les suivantes :

- expérience aléatoire ;
- expérience naturelle ;
- expérience quasi aléatoire ;
- régression discontinue ;
- méthode d'appariement ;
- variables instrumentales ;
- double différence ;
- régression.

Dans la mesure où, d'une part, chaque méthode peut faire l'objet de plusieurs dénominations et déclinaisons en anglais, et, d'autre part, notre recherche se fait sur la base d'expressions

⁴ Certains articles sans traduction ont néanmoins été relevés, mais leur nombre restait très marginal pour notre étude.

⁵ La classification JEL est une méthode standardisée pour classer les travaux académiques en science économique. Elle a été développée par le *Journal of Economic Literature*. Voir <https://www.aeaweb.org/econlit/jelCodes.php> pour une description des codes.

⁶ Il aurait été possible d'utiliser les mots clés renseignés dans la base, mais cet exercice se révélait plus délicat dans la mesure où ils ne sont pas harmonisés.

⁷ Il convient de souligner que les évaluations qui portent sur un pays donné ne sont pas nécessairement effectuées par des chercheurs de ce pays. Les résultats s'interprètent donc non pas comme l'activité du milieu académique de ce pays mais plutôt comme l'intérêt que portent les scientifiques nationaux et étrangers à l'évaluation d'impact des politiques publiques d'un pays donné.

exactes obligeant à prendre en compte le pluriel et le singulier, au total 49 expressions clés ont été utilisées pour retracer les huit méthodes. L'ensemble de ces expressions est présenté dans le Tableau 1 ci-après.

Par ailleurs, ce choix d'expressions est aussi la résultante d'un travail de tests successifs. Nous avons en effet conduit des recherches sur des échantillons de revues françaises⁸ et anglo-saxonnes⁹ afin d'identifier les différentes formulations possibles pour les méthodes d'évaluation d'impact ainsi que les expressions générales les plus fréquemment utilisées dans les résumés.

De ce premier travail d'identification, nous obtenons 13 050¹⁰ articles dans la base EconLit. Nous exportons l'ensemble des résultats dans le logiciel Stata© pour les traitements ultérieurs.

⁸ *Économie et Statistique, Économie et prévision, Annales d'économie et de statistique.*

⁹ *The Quarterly Journal of Economics, Journal of Public Economics, The American Economic Review, Journal of Labor Economics, Journal of Econometrics, Journal of Development Economics.*

¹⁰ L'extraction a eu lieu le 15 mai 2018.

Tableau 1 – Liste des expressions pour construire la requête de recherche

Expérience aléatoire	Expérience naturelle	Expérience quasi aléatoire
randomized trial randomized trials randomized evaluation randomized evaluations randomized experiment randomized experiments randomized controlled trial randomized controlled trials RCT RCTs randomised trial randomised trials randomised evaluation randomised evaluations randomised experiment randomised experiments randomised controlled trial randomised controlled trials	field experiment field experiments natural experiment natural experiments regional experiment regional experiments	quasi-experiment quasi-experimental study quasi-experimental method quasi-randomly assigned
Régression discontinue	Méthode d'appariement	Variables instrumentales
regression discontinuity discontinuity analysis	matched comparisons matched-comparison design matching propensity score	instrumental variable instrumental variables
Double différence	Régression	Autres
difference-in-differences differences-in-differences double difference double differences triple difference triple differences	regression approach	causal impact causal impacts causal effect causal effects counterfactual treatment effect treatment effects

Note de lecture : Pour chaque méthode, il existe plusieurs dénominations possibles à la fois au singulier et au pluriel. Ce tableau récapitule, pour chaque méthode et pour la catégorie générale, les expressions recherchées dans la base EconLit de manière exacte. Nous avons aussi tenu compte des différences existantes entre l'anglais américain et l'anglais britannique.

Source : Auteurs

1.3. Création de la base de travail

Malgré le soin apporté dans la définition des termes de la requête, tous les articles extraits de la base EconLit ne correspondent pas à des travaux d'évaluations d'impact *ex post*. Il existe en effet, dans la base constituée, un grand nombre d'articles couvrant d'autres types de recherche en économie. Le principal filtre utilisé pour affiner notre base est le champ « pays ». En effet, sans identificateur de pays, l'article de recherche correspond le plus souvent à une publication théorique ou de réflexion, non adossée à une recherche empirique conduite sur un dispositif ou une politique publique mise en œuvre dans un pays ou un groupe de pays, donc en dehors de notre champ d'étude. Nous supprimons ainsi 2 396 observations de la base qui se compose alors de 10 654 articles.

Nous avons aussi constaté que lorsque le champ pays comportait plusieurs entrées (pays ou groupe de pays), l'article de recherche était bien souvent un article de recherche en macroéconomie, ce qui est en dehors de notre champ d'étude. C'est pourquoi nous avons fait le choix de supprimer 1 986 articles comportant plusieurs entrées dans le champ pays. En supprimant ces observations, nous obtenons une base de travail constituée de 8 668 articles.

La suppression des articles avec plusieurs pays est surtout problématique pour un nombre marginal d'évaluations d'impact qui vont utiliser des effets de bord des deux côtés d'une frontière. Ce cas peut se produire au plus 350 fois dans nos observations supprimées (correspondant au nombre d'articles portant sur deux pays). Cette possibilité d'erreur reste acceptable comparée à la taille de notre échantillon total (elle représente au maximum 5 % des articles).

Enfin, par souci de disposer uniquement des années complètes, nous avons supprimé les entrées correspondant à l'année 2017, incomplète au moment du démarrage de nos travaux, soit 1 204 articles.

Au total, ces différentes opérations permettent d'aboutir à une base de données finale composée de 7 464 articles de recherche portant sur des évaluations d'impact *ex post*.

Malgré le soin apporté à la définition des termes de la requête, certaines erreurs peuvent exister dans la base constituée, à la fois parce que des articles identifiés peuvent ne pas correspondre à notre cible et parce que des articles s'inscrivant dans notre cible peuvent échapper à notre recherche. Pour mieux comprendre ces erreurs, nous avons réalisé deux tests pour la France seulement. Le premier test a consisté à analyser l'ensemble des articles de la base portant spécifiquement sur la France. Le second test a passé en revue l'ensemble des articles publiés dans neuf revues françaises présentes dans la base EconLit depuis la date de leur versement dans celle-ci et jusqu'en 2016 (voir annexe F).

Cet examen de la base pour la France indique que les erreurs de ciblage existent dans les deux sens. Sur les 180 articles identifiés, une trentaine d'articles ne correspondent pas à notre ciblage (principalement des articles en économie structurelle et comportementale, des articles de synthèse et quelques articles descriptifs). Par ailleurs, une cinquantaine d'articles publiés dans des revues françaises échapperaient à notre ciblage. Ces derniers ne sont pas identifiés en raison principalement de l'absence d'indications méthodologiques dans le résumé de ces articles, ou de l'absence d'indicateur du pays dans EconLit.

Néanmoins, ces erreurs de ciblage ne sont pas spécifiques à la France et peuvent potentiellement affecter les autres pays de notre étude. En particulier, les pays non anglo-saxons peuvent être affectés comme la France par un biais de sous-estimation pour les articles parus dans leurs revues nationales. En revanche, pour les pays anglo-saxons, on peut s'attendre à une surestimation du nombre d'articles du fait du premier type de biais (existence de faux positifs). En effet, les méthodes seraient moins indiquées dans le résumé des articles des revues nationales car les standards seraient moins exigeants que pour les revues anglo-saxonnes. Si nous avons choisi de rechercher les mots clés directement dans le corps du texte, il aurait fallu recourir à une traduction, dans chaque langue nationale, des expressions désignant les méthodes. Cela n'aurait pas nécessairement permis de résoudre ce problème en raison de la difficulté de recenser toutes les variantes possibles de traduction. De plus, il aurait été impossible de savoir si l'article en lui-même mobilise ces méthodes ou bien fait référence à d'autres travaux les utilisant.

Nous avons réalisé des tests de sensibilité de nos classements à ces biais à partir de l'analyse de la base France. La méthode et les résultats de ces tests sont présentés de manière détaillée dans l'annexe F. Compte tenu des différences dans les volumes d'articles identifiés dans notre base entre les pays, il apparaît peu vraisemblable que ces erreurs modifient *in fine* les classements réalisés ou l'évolution temporelle de la production, comme les résultats de ces tests de sensibilité l'indiquent.

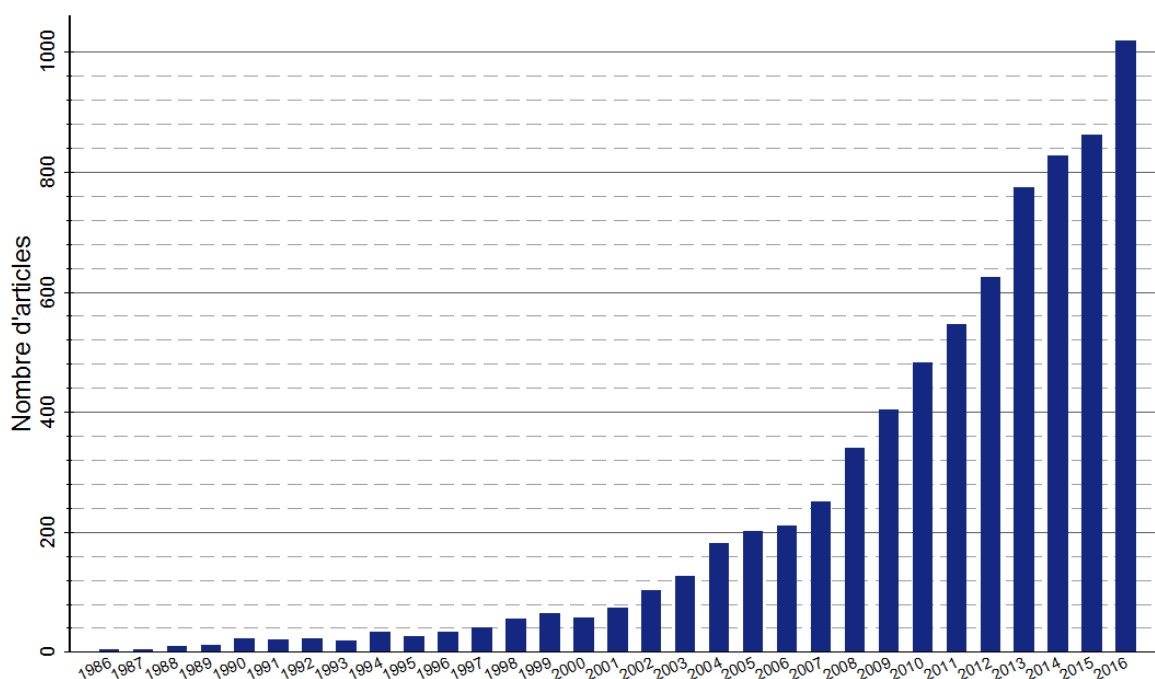
2. Les résultats

Avant d'étudier plus spécifiquement les sept pays retenus, plusieurs analyses peuvent être réalisées à partir des résultats de volumétrie générale sur l'ensemble de la base. Nous en tirons les résultats généraux présentés ci-après.

2.1. Près de 7 500 articles d'évaluation d'impact pour tous les pays

Le volume total d'articles scientifiques d'évaluation *ex post* produit sur l'ensemble des pays s'élève, selon notre technique de recherche, à 7 464. La grande majorité de ces articles a été publiée après 1986. À partir de cette date, le nombre d'articles publiés correspondant à ce type d'évaluation n'a cessé d'augmenter année après année (Figure 1). C'est cependant au cours des années 2000 que l'on observe un net infléchissement à la hausse de la production. L'année 2016 poursuit et amplifie cette tendance et apparaît comme une année prolifique pour les évaluations d'impact avec plus de 1 000 articles recensés pour cette seule année.

Figure 1 – Nombre total d'articles présentant des résultats d'évaluations d'impact publiés chaque année



Note de lecture : Il n'existe que très peu d'articles avant 1986.

Source : EconLit, traitement des auteurs

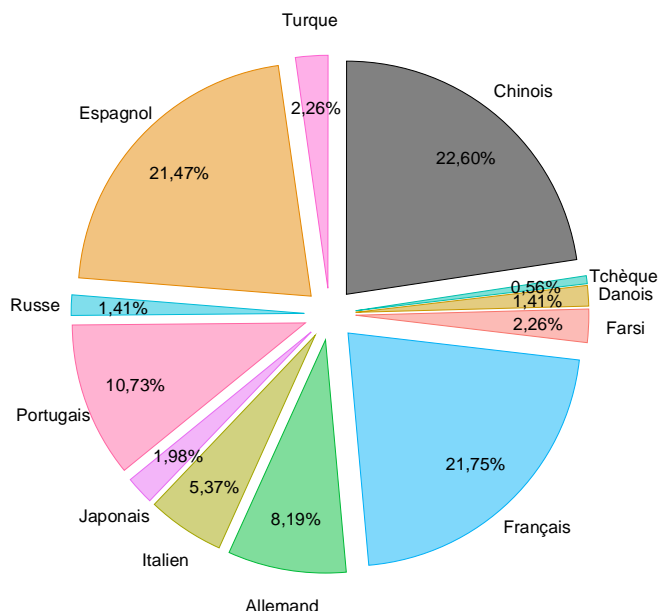
Plus de 95 % des articles de notre base d'étude sont rédigés en anglais, soit 7 107 articles (Table 1 et Figure 2). Le chinois, le français et l'espagnol sont les trois autres langues les plus usitées avec respectivement 80, 77 et 76 articles disponibles. Elles représentent chacune environ 22 % des articles rédigés dans une langue autre que l'anglais. L'anglais est

ainsi pour les évaluations d'impact, comme pour la recherche économique dans son ensemble, la langue dominante utilisée dans les articles.

Table 1 : Langue des articles

	Fréquence	Pourcentage
Allemand	29	0,39
Anglais	7107	95,22
Chinois	80	1,07
Croate	1	0,01
Danois	5	0,07
Espagnol	76	1,02
Farsi	8	0,11
Français	77	1,03
Italien	19	0,25
Japonais	7	0,09
Portugais	38	0,51
Russe	5	0,07
Slovaque	1	0,01
Tchèque	2	0,03
Turque	8	0,11
Ukrainien	1	0,01
<i>N</i>	7464	

Figure 2 – Répartition des langues hormis l'anglais



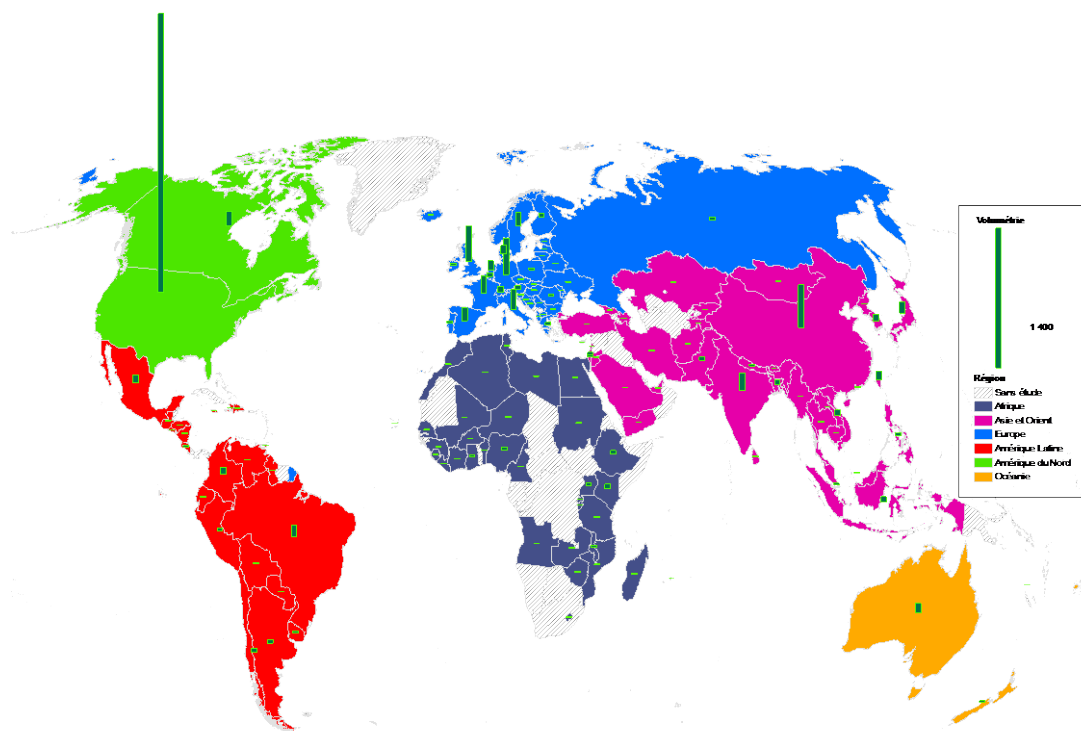
Source : EconLit, traitement des auteurs

2.1.1. Une production mondiale dominée par les États-Unis

La carte de la Figure 3 permet de visualiser l'ensemble des pays qui comptent au moins une évaluation les concernant jusqu'en 2016. La volumétrie est représentée sous forme d'histogrammes associés à chaque pays, tandis que les pays ne comptant aucune étude sont indiqués en hachuré sur la carte.

Les États-Unis apparaissent nettement comme le leader mondial de la production des évaluations d'impact. La Grande-Bretagne, l'Allemagne et l'Italie constituent, dans l'ordre, le top 3 des pays européens. En Asie, la Chine et l'Inde sont les deux pays les plus représentés, loin devant les autres pays du continent (Indonésie, Singapour), tandis qu'en Afrique aucun pays ne se caractérise par une volumétrie importante d'évaluations. Parmi les pays ne totalisant aucune étude dans notre base de travail, peuvent être mentionnés l'Afrique du Sud et des pays du Proche-Orient (Israël, Liban) qui disposent des ressources académiques ainsi que des structures politico-administratives rendant pourtant possibles et appropriés ce type de travaux.

Figure 3 – Nombre d'articles d'évaluation d'impact des pays et zonage en grandes régions

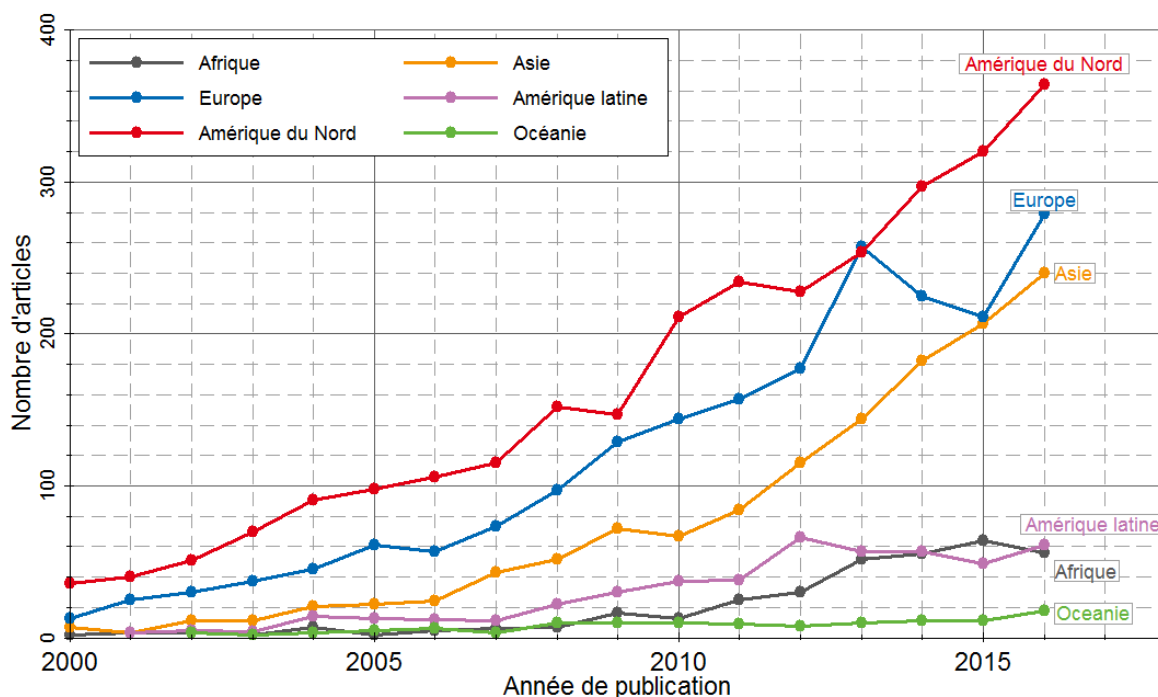


Note de lecture : La hauteur des histogrammes correspond au nombre total d'articles par pays dans la base d'étude.

Source : Traitement des auteurs, zonage défini par EconLit

Nous présentons aussi la typologie en grandes régions agrégées sur la base du zonage disponible dans EconLit. L'évolution annuelle par région (Figure 4) montre que les évaluations d'impact étaient relativement peu nombreuses au début des années 2000 et n'ont cessé d'augmenter au fil des ans jusqu'en 2016. Ce sont les deux pays d'Amérique du Nord, Canada et États-Unis, qui cumulent chaque année le plus grand nombre d'études les concernant (principalement les États-Unis). La seconde grande région est l'Europe, rattrapée durant les dernières années par l'Asie. En termes de dynamique, nous constatons une accélération de la production aux alentours de l'année 2005 pour les pays de l'Ouest (Amérique du Nord et Europe) alors que l'Asie connaît une accélération cinq à six ans plus tard. Même si pour l'Afrique le nombre d'évaluations a triplé entre 2010 et 2015, le continent semble connaître une stabilisation depuis lors. L'Amérique latine suit sensiblement la même trajectoire que l'Afrique avec des volumes d'études comparables.

Figure 4 – Nombre d'articles d'évaluation d'impact par année et par grande région entre 2000 et 2016



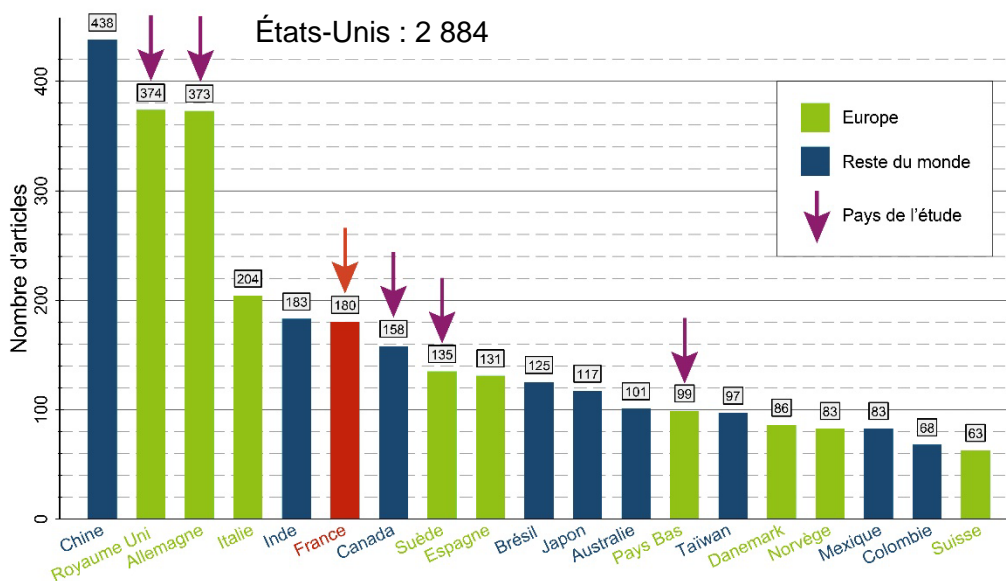
Note de lecture : Chaque point d'une courbe représente le nombre d'articles publiés qui concernent les pays appartenant à la région mentionnée. L'Amérique du Nord s'entend seulement avec le Canada et les États-Unis.

Source : EconLit, traitement des auteurs

La Figure 5 détaille plus précisément la volumétrie pour les vingt pays concernés par le plus grand nombre d'études. Les États-Unis présentent de loin le plus grand volume d'études (2 884). Ils sont suivis par la Chine, mais qui en compte près de sept fois moins (438). Le Royaume-Uni et l'Allemagne présentent le même volume d'études (respectivement 374 et 373), loin devant la France avec seulement 180 études répertoriées. L'Inde et l'Italie se placent devant la France (respectivement 204 et 183) et le Canada juste derrière (158). Les pays du nord de l'Europe (Suède, Danemark, Norvège) sont présents, bien qu'un peu plus loin dans le classement.

Si l'on s'intéresse au nombre de publications rapporté à la population en 2014, toujours pour les vingt pays ayant le plus d'études, la hiérarchie présentée précédemment est significativement modifiée (Figure 6). Les pays du nord de l'Europe sont de loin ceux qui bénéficient du plus grand nombre d'évaluations par habitant (Norvège, Danemark, Suède). Viennent ensuite les États-Unis, la Suisse, les Pays-Bas, le Royaume-Uni et l'Allemagne. La France, avec 2,72 études par million d'habitants, est près de deux fois moins pourvue que ses voisins allemand (4,63 études par million d'habitants) et britannique (5,81 études par million d'habitants). Relativement à sa population, elle produit également moins d'évaluations d'impact que le Canada, l'Australie ou l'Italie. Elle est en revanche proche mais légèrement en dessous de l'Espagne.

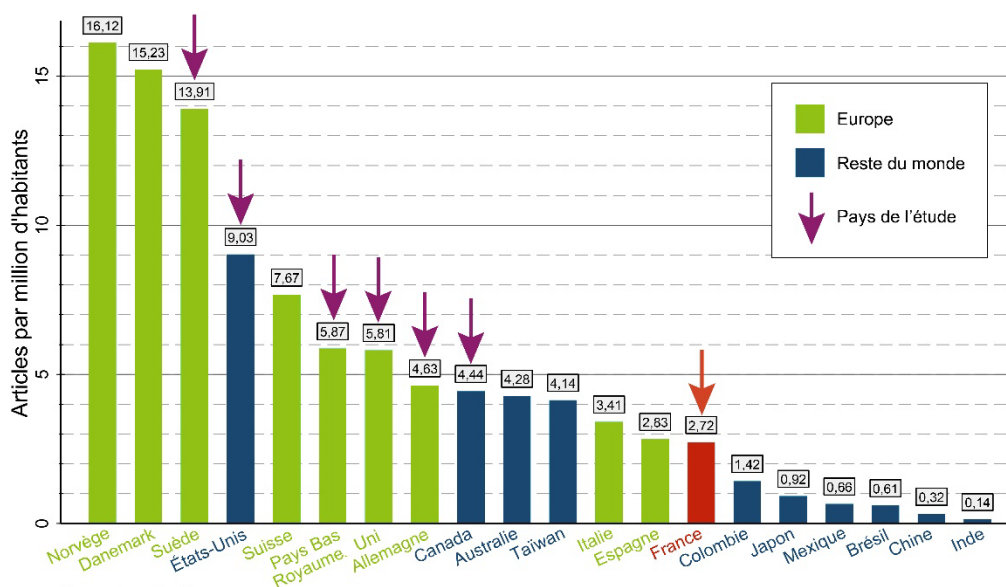
Figure 5 – Nombre total d'articles d'évaluation d'impact par pays pour le top 20



Note de lecture : On reporte les vingt premiers pays en fonction du nombre d'évaluations d'impact qui leur est associé jusqu'en 2016. La valeur pour les États-Unis est reportée séparément pour une meilleure lecture du graphique.

Source : EconLit, traitement des auteurs

Figure 6 – Nombre d'articles d'évaluation d'impact par million d'habitants pour le top 20



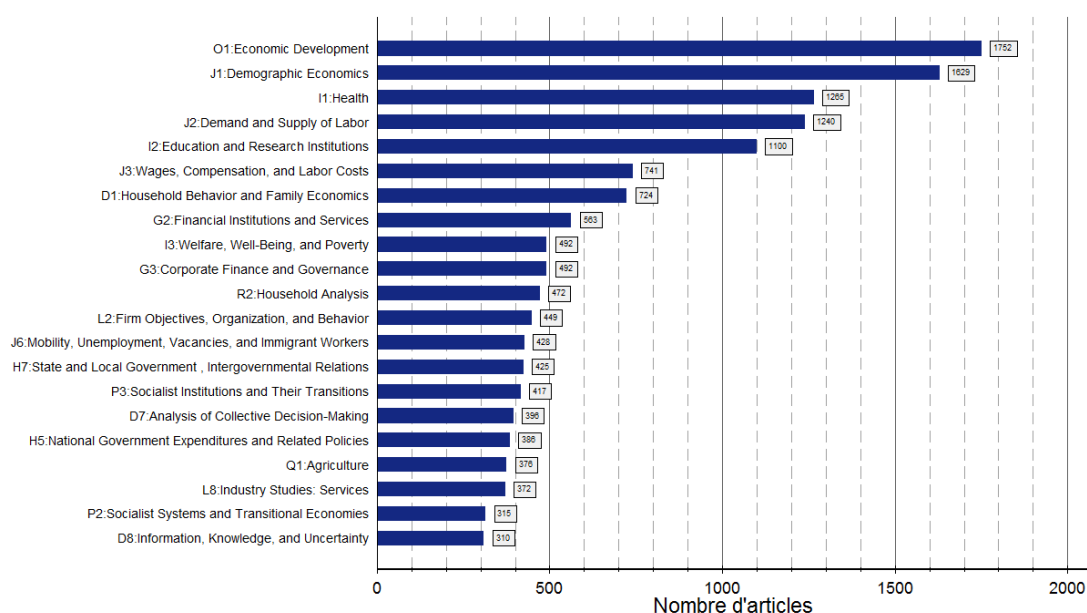
Note de lecture : Ce graphique exprime, pour les vingt pays ayant produit le plus d'évaluations d'impact jusqu'en 2016, le nombre d'articles divisé par la population du pays en 2014 (en million d'habitants).

Source : EconLit, Penn World Table V9.0 pour la population en 2014, traitement des auteurs

2.1.2. Des évaluations portées par l'économie du développement et centrées sur les secteurs de l'emploi, de la santé et de l'éducation

Selon la classification en codes JEL¹¹, l'économie du développement (O1) et l'économie démographique (J1) sont les deux domaines les plus représentés (Figure 7 et Figure 17 en annexe A). La catégorie J1, économie démographique, correspond à l'économie de la famille et de la population : elle englobe les politiques publiques liées à la famille, aux personnes âgées et handicapées, les discriminations (hors de l'emploi), l'économie du genre et des minorités et l'immigration. Les trois autres grands domaines sont la santé, le travail et l'éducation. Il existe un écart important entre les cinq premières catégories et les autres. En dynamique, on note que chaque catégorie progresse sur la période, et en particulier l'économie du développement à partir des années 2010 (Figure 8 et Figure 18 en annexe A).

Figure 7 – Codes JEL de catégorie 2 les plus représentés

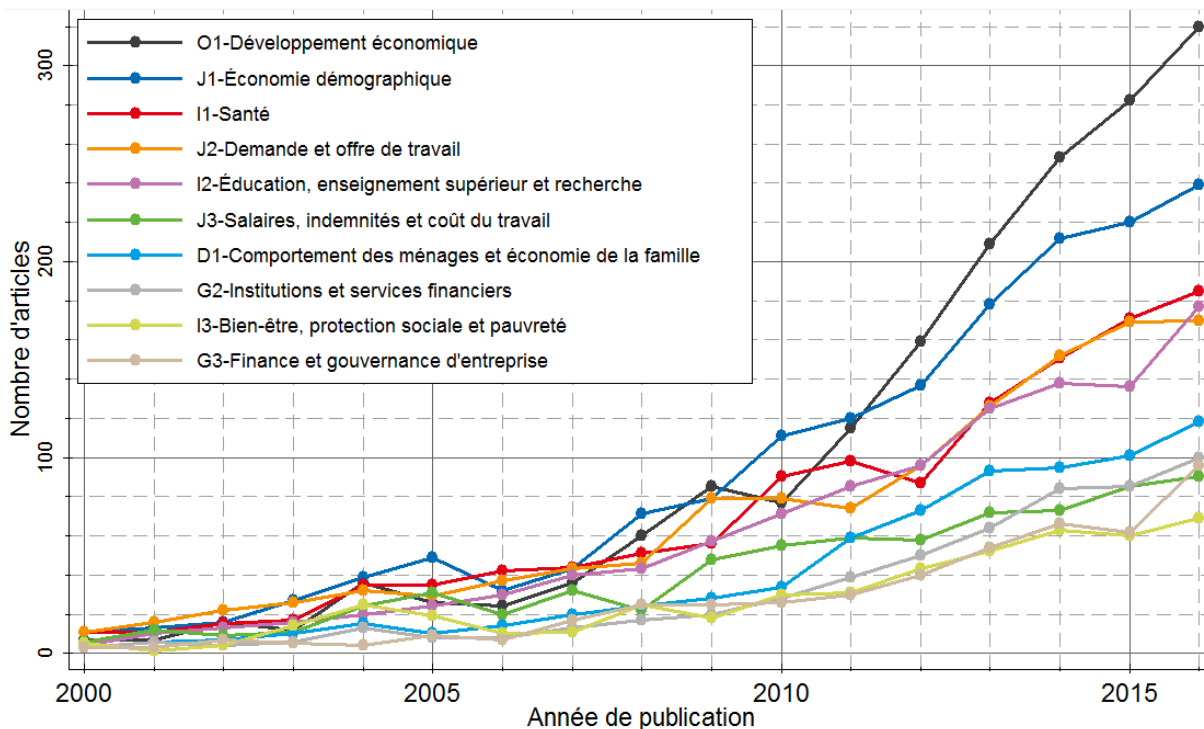


Note de lecture : Chaque somme correspond au nombre de fois, de manière unique, qu'un code JEL intermédiaire apparaît. Par exemple pour un article avec les codes JEL : J10, J11, J31, on compte 1 unité pour J1 et une unité pour J3. Nous conservons les dénominations en langue anglaise pour éviter des interprétations de traduction.

Source : EconLit, traitement des auteurs

¹¹ Les codes JEL permettent une analyse en trois niveaux de détail. Les codes JEL sont constitués d'une lettre, qui correspond à la catégorie de niveau 1, la plus large, suivie de deux chiffres. Le premier chiffre correspond à une catégorie intermédiaire et le dernier chiffre à la catégorie la plus fine. Chaque article peut avoir un ou plusieurs codes JEL (81 articles seulement de notre base ne possèdent pas de code JEL).

Figure 8 – Évolution des 10 premiers codes JEL de catégorie 2 entre 2000 et 2016

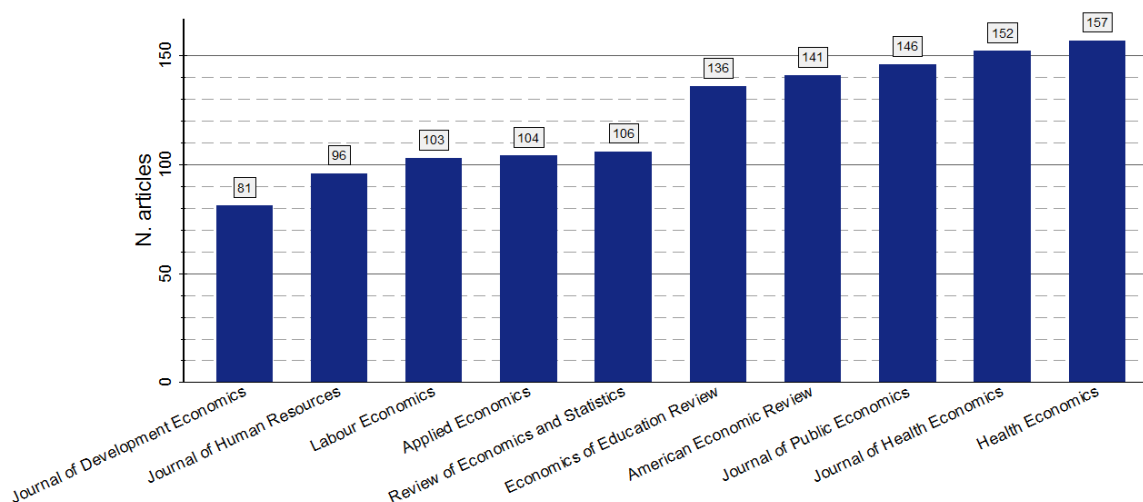


Source : EconLit, traitement des auteurs

2.1.3. Des articles d'évaluation publiés dans des revues généralistes en économie

La Figure 9 indique le classement des dix premiers journaux académiques qui publient le plus d'évaluations d'impact (voir également Figure 19 en annexe B pour le classement des journaux au-delà de la dixième place). Il comprend des journaux spécialisés dans les domaines mis en évidence précédemment (santé, éducation, travail, économie du développement). Les deux journaux qui publient le plus d'évaluations d'impact sont d'ailleurs spécialisés dans le domaine de la santé : *Health Economics* et *Journal of Health Economics*. Mais il comprend aussi des revues généralistes en économie, dont certaines sont de tout premier plan du point de vue académique (*Journal of Public Economics*, *American Economic Review*, *Review of Economics and Statistics*).

Figure 9 – Top 10 des journaux qui publient le plus d'évaluations d'impact



Note de lecture : Nombre d'articles par journal sur l'ensemble de la base.

Source : EconLit, traitement des auteurs

2.2. Par rapport aux pays les plus avancés, une production française relativement modeste, mais en rattrapage

Dans cette partie, nous limitons notre analyse à un échantillon de sept pays listés ci-après :

- Allemagne ;
- Canada ;
- États-Unis ;
- France ;
- Pays-Bas ;
- Royaume-Uni ;
- Suède.

Le choix de cet échantillon répond à une double préoccupation méthodologique :

- d'une part, le fait de pouvoir situer la France par rapport à des pays au niveau de vie comparable et disposant à la fois de structures académiques et d'un système politico-administratif proches ;
- d'autre part, le choix de comparer la France à des pays réputés comme les plus avancés en matière d'évaluation des politiques publiques dans la littérature internationale¹².

¹² Cette sélection s'est appuyée sur l'analyse de plusieurs études portant sur les systèmes d'évaluation nationaux de divers pays (Bachtler *et al.*, 2000 ; Furubo, Rist, Sandahl, 2002 ; Jacob, Varone, 2004 ; Polverari, Bachtler, 2004 ; Sharma *et al.*, 2013 ; Lázaro, 2015 ; OCDE, 2015 ; SGI network, 2017). Pour chacune de ces études, nous avons recensé les pays considérés comme les plus avancés en matière d'évaluation selon les critères propres à l'étude (« culture » d'évaluation, institutionnalisation, maturité de la pratique, mise en œuvre effective, etc.). Nous avons ensuite retenu, dans l'ordre, les pays les plus cités dans l'ensemble des études.

Pour des questions de faisabilité et dans la mesure où l'étude bibliométrique sera complétée par une analyse monographique approfondie, nous avons fait le choix de limiter l'échantillon à sept pays. Cela explique que n'ont pas été intégrés dans notre analyse plusieurs pays (notamment l'Australie, la Suisse, l'Espagne ou encore l'Italie) jugés initialement pertinents pour notre exercice de comparaison.

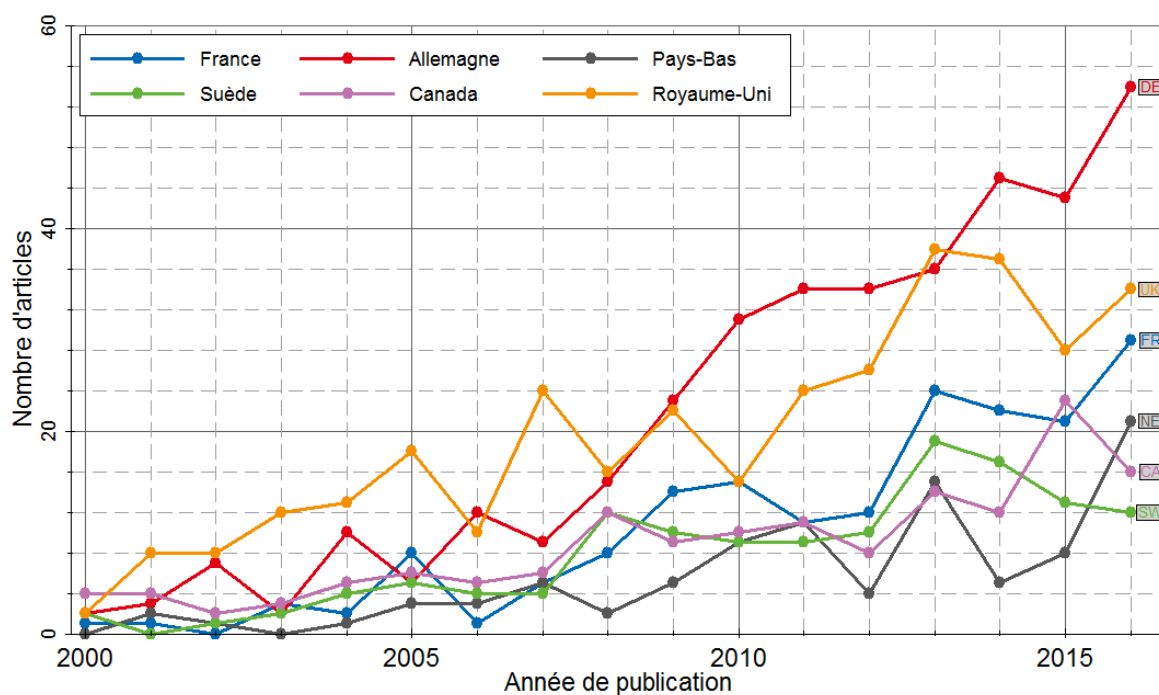
2.2.1. Une dynamique positive dans tous les pays de l'échantillon, en particulier en France

Le graphique de la Figure 10 présente l'évolution annuelle du nombre d'évaluations d'impact portant sur les pays de notre échantillon hors États-Unis. Sur la période 2000-2016, si l'on met à part les États-Unis, le Royaume-Uni a été le premier pays du reste de l'échantillon à produire des évaluations d'impact de façon régulière et en progression durant toute la première partie des années 2000. Le Royaume-Uni est ensuite rattrapé par l'Allemagne dans la seconde partie des années 2000. De son côté, la France a commencé à produire des études plus tardivement que ses voisins allemand et britannique sans pour autant arriver à les égaler, bien qu'elle se rapproche du niveau du Royaume-Uni en fin de période avec une trentaine d'articles publiés en 2016.

Pour chaque pays, la tendance générale est à une augmentation de la production d'évaluations d'impact sur la période. Enfin, notons que la production sur les États-Unis est sans commune mesure avec celle des autres pays de l'échantillon. En dépit de ce volume initial important, le flux annuel d'évaluations pour les États-Unis continue à croître sur toute la période (voir la Figure 20 en annexe C pour le même graphique avec les États-Unis).

Lorsque l'on rapporte l'évolution du nombre de publications à celui de la population (Figure 21 en annexe D), la France se place quasiment chaque année en dernière position parmi les pays étudiés. Malgré une population importante, les États-Unis sont en deuxième position juste derrière la Suède dans notre sélection de pays. Notons que tous les pays de notre échantillon d'étude connaissent une progression du nombre d'articles rapporté à la population.

Figure 10 – Évolution du nombre d'articles d'évaluation d'impact publiés pour les pays de l'échantillon sauf les États-Unis entre 2000 et 2016



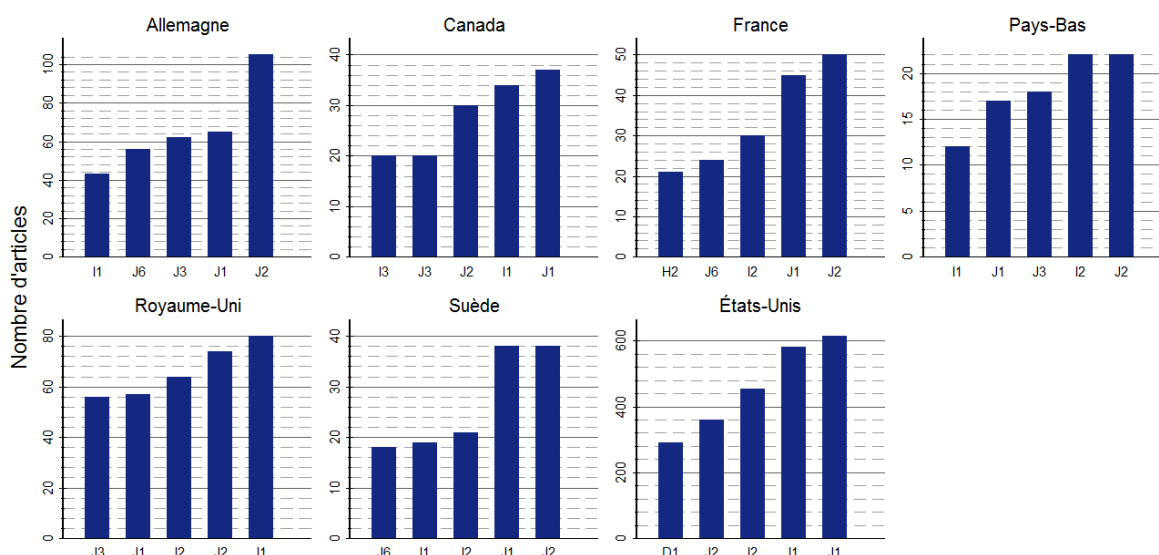
Note de lecture : Nous ne reportons pas les États-Unis sur ce graphique pour faciliter la lecture. Le graphique avec les États-Unis est disponible en annexe C.

Source : EconLit, traitement des auteurs

2.2.2. Des thématiques de recherche communes : emploi, santé, éducation

En ce qui concerne les domaines d'action publique les plus évalués, la Figure 11 permet d'observer une relative homogénéité entre les pays. Pour tous les pays de notre échantillon, les domaines de l'emploi, de la santé, de l'éducation et de l'économie de la famille et de la population (économie démographique) sont les plus représentés. Pour quatre des sept pays (Allemagne, France, Pays-Bas et Suède), le code JEL J2, relatif à l'offre et à la demande de travail, est le premier domaine investi. En France, après les questions relatives au marché du travail, ce sont les thématiques de la famille et de la population qui apparaissent comme les plus traitées. Le Royaume-Uni se caractérise par la prépondérance du champ « santé » (I1), et les États-Unis et le Canada par celui de l'économie démographique.

Figure 11 – Top 5 codes JEL par pays



Note de lecture : Nombre d'apparitions d'un code JEL de catégorie deux par pays jusqu'en 2016. Attention, l'échelle des ordonnées change pour chaque pays.

Liste des codes JEL : **D1**-Comportement des ménages et économie de la famille ; **I1**-Santé ; **I2**-Éducation, enseignement supérieur et recherche ; **I3**-Bien-être, protection sociale et pauvreté ; **J1**-Économie démographique ; **J2**-Demande et offre de travail ; **J3**-Salaires, indemnités et coût du travail ; **J6**-Mobilité, chômage, postes vacants et travailleurs immigrés.

Source : EconLit, traitement des auteurs

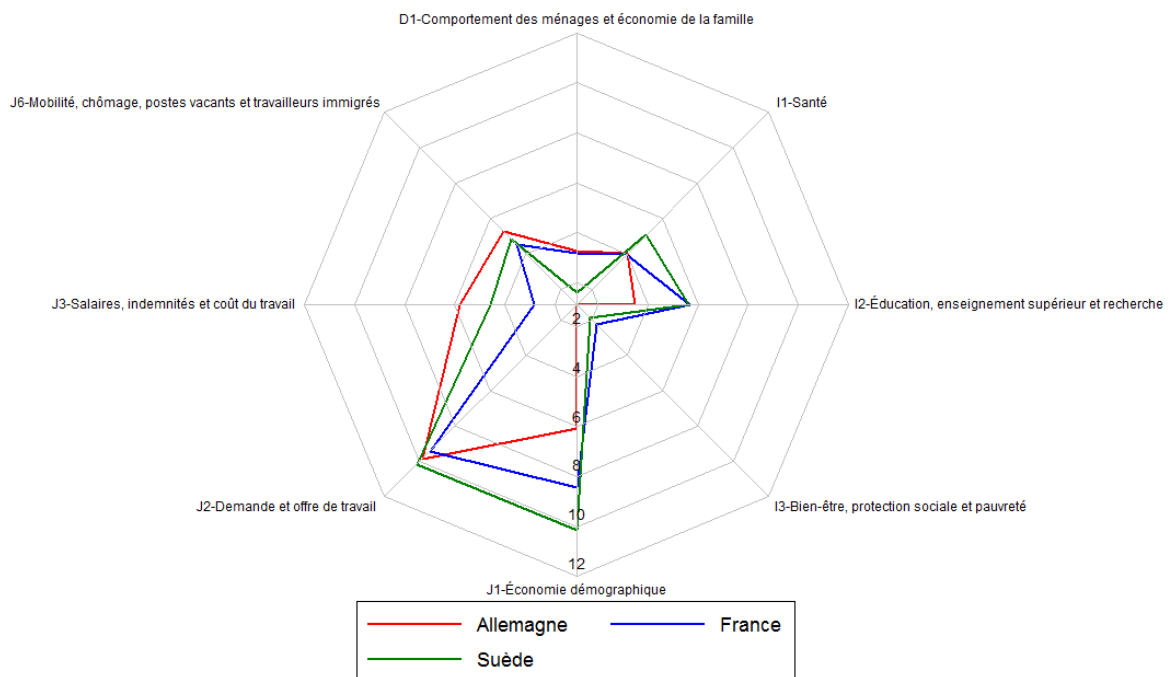
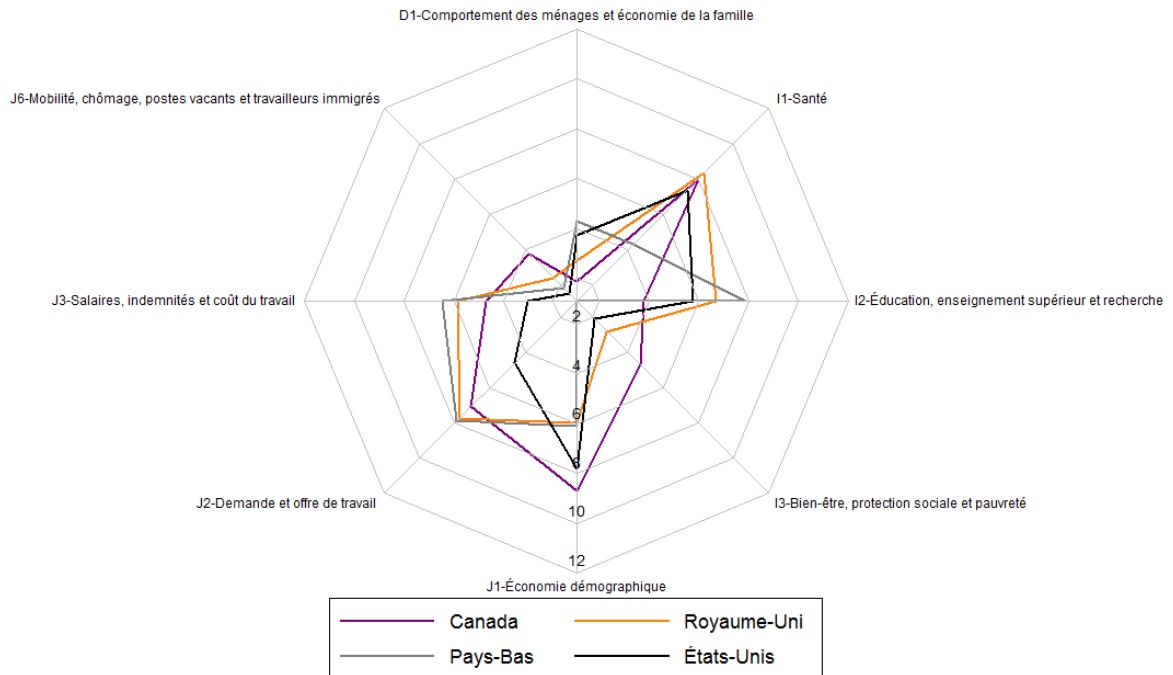
L'existence d'un nombre réduit de domaines d'action prépondérants dans les pays de notre échantillon (au total, seuls 8 codes JEL différents sont présents dans les top 5 de tous les pays) amène à se questionner sur l'existence de « profils » d'évaluation plus ou moins proches entre ces pays. Nous représentons ces profils sous forme de graphique de type « radar », en calculant pour chaque pays le pourcentage d'occurrences des codes JEL les plus représentés parmi le nombre total de codes JEL cités dans les études portant sur le pays (Figure 12). Schématiquement, deux grandes catégories de pays émergent : d'une part, un premier groupe (Canada, Royaume-Uni, Pays-Bas, États-Unis) orienté vers les

thématiques de santé et d'éducation ; d'autre part, un second groupe (Allemagne, France, Suède) nettement plus préoccupé par les questions relatives au marché du travail. Au sein du second groupe, l'Allemagne et la Suède présentent les profils les plus extrêmes, puisque si l'on additionne les scores des trois codes JEL relatifs au marché du travail (J2, J3, J6), la thématique de l'emploi représente plus de 20 % des codes JEL dans ces pays.

Une autre manière de considérer les « profils » nationaux est de calculer la distance entre les scores (en pourcentage du total) de l'ensemble des codes JEL de catégorie 2 dans chaque pays. La Figure 13 présente l'éloignement des pays de notre échantillon vis-à-vis de la France selon cette métrique. Elle confirme que l'Allemagne et la Suède sont les pays les plus proches de la France en termes de couverture sectorielle des évaluations, tandis que les autres pays de l'échantillon présentent un profil différent, en particulier les États-Unis. Le classement des pays recoupe en fait la distinction classique entre les pays anglo-saxons ou proches de par leur tradition politique et administrative ou leur régime (États-Unis, Royaume-Uni, Canada, Pays-Bas) et les pays européens continentaux (France, Allemagne, Suède). Les Pays-Bas mis à part, ce découpage témoigne aussi de l'influence exercée par le modèle social national ou le régime d'État-providence sur les préoccupations de politiques publiques et, par conséquent, les évaluations d'impact qui y sont conduites¹³.

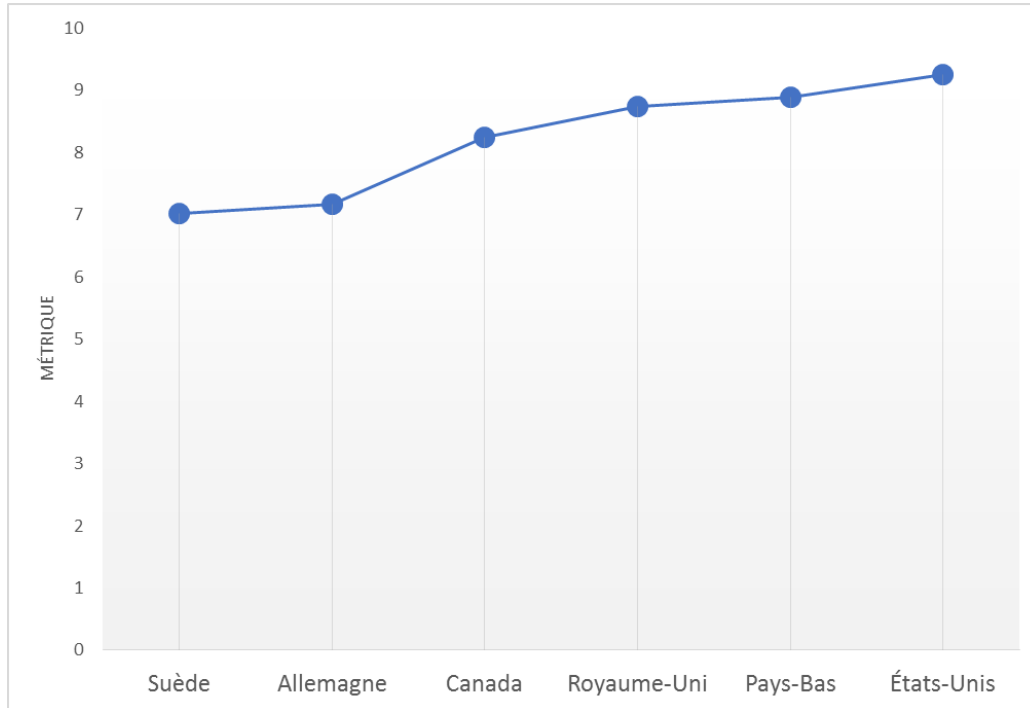
¹³ Voir notamment la catégorisation classique des États-providences proposée par G. Esping-Andersen (1999).

Figure 12 – Pourcentage des principaux codes JEL par pays



Note de lecture : Pour chaque code JEL retenu et par pays, on calcule le pourcentage d'occurrences du code JEL par rapport au nombre total de codes JEL dans le pays.

Source : EconLit, traitement des auteurs

Figure 13 – Distance des pays à la France en termes de scores des codes JEL

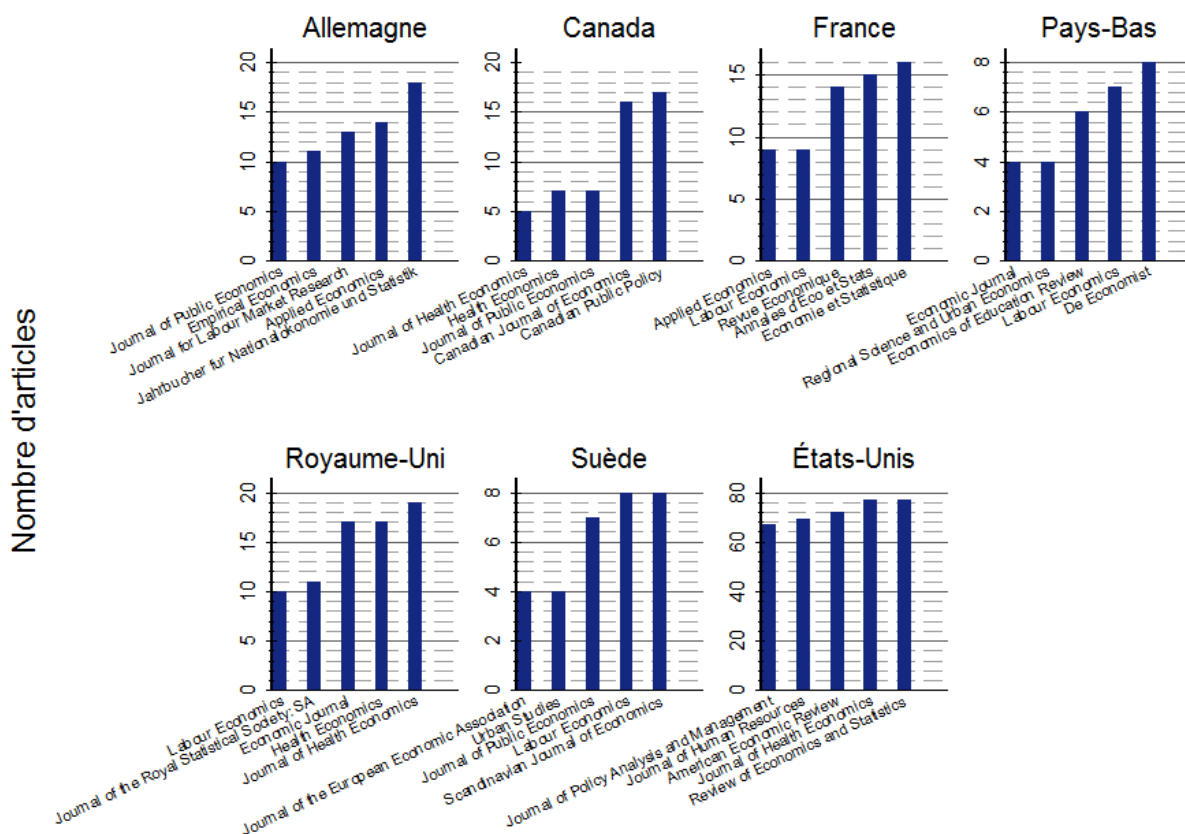
Note de lecture : La distance est calculée comme la racine carrée de la somme des différences des carrés des scores des codes JEL entre un pays et la France. Le score d'un code JEL est son pourcentage d'apparitions dans le total des occurrences des codes JEL dans un pays. La distance se mesure donc en points de pourcentage.

Source : EconLit, traitement des auteurs

2.2.3. Des évaluations publiées majoritairement dans les revues nationales (hors monde anglo-saxon)

S'agissant des journaux de publication des articles, les sept pays de notre échantillon se distinguent cette fois nettement les uns des autres, comme le montre la Figure 14. En effet, pour chacun des pays non anglo-saxons, c'est une revue nationale ou régionale qui publie le plus d'évaluations d'impact (*Jahrbücher für Nationalökonomie und Statistik* en Allemagne, *Économie et Statistique* en France, *De Economist* aux Pays-Bas, *The Scandinavian Journal of Economics* en Suède). Cependant, par rapport à d'autres pays de l'échantillon (Suède, Pays-Bas, Allemagne), les articles portant sur la France et le Canada sont de façon plus marquée publiés dans les revues nationales, ce qui peut leur conférer une moindre visibilité internationale (même si pour le Canada les deux premières revues publiant le plus d'évaluations d'impact, nationales, sont en langue anglaise). La prépondérance de certaines revues dans les pays illustre également la spécification thématique des évaluations d'impact dans ces pays : par exemple, pour le Royaume-Uni figurent en tête deux revues spécialisées dans le domaine de la santé, le *Journal of Health Economics* et *Health Economics*. Ainsi, le classement des revues révèle les spécificités nationales dans les domaines traités (santé, travail, etc.) et traduit le fait que l'ancrage national des évaluations d'impact est important – les évaluations portant sur un pays sont fréquemment publiées dans les revues nationales associées, en particulier pour la France et le Canada et dans une moindre mesure pour les Pays-Bas, l'Allemagne et la Suède.

Figure 14 – Top 5 des revues par pays



Note de lecture : Nombre d'articles concernant un pays jusqu'en 2016 inclus. Attention, les échelles des ordonnées entre les pays ne correspondent pas.

Source : EconLit, traitement des auteurs

2.2.4. Les grands acteurs académiques nationaux très présents sur le champ de l'évaluation

Pour terminer, nous effectuons un classement des institutions les plus productrices d'évaluations d'impact sur chacun des pays de notre échantillon. Ce classement n'a pas vocation à refléter la qualité du travail produit par ces institutions, mais à comparer, à partir de leur volume de production, l'intérêt porté aux évaluations d'impact par les différentes institutions de recherche. Pour cette raison, nous utilisons un système de comptabilisation des institutions qui ne tient pas compte du nombre de pages, du nombre de citations obtenues ou de la qualité de la revue dans laquelle l'article a été publié. Par ailleurs, contrairement à la méthodologie choisie, par exemple, par Combes et Linnemer (2003)¹⁴, nous utilisons les affiliations des auteurs des articles telles qu'elles sont déclarées dans la base EconLit afin d'attribuer à chaque institution les travaux qu'elle a effectivement soutenus au cours du temps.

¹⁴ Cette étude n'utilise pas les affectations renseignées dans la base EconLit mais recherche l'affectation actuelle des chercheurs à partir des sites internet des centres de recherche. Toute la production passée des chercheurs est attribuée à leur institution actuelle, ce qui fournit une mesure du « capital humain » disponible dans l'institution.

Cependant, recourir aux affiliations déclarées dans le champ adresse de la base EconLit reste une entreprise difficile tant les pratiques en termes de signatures peuvent varier en fonction des pays, du type d'institution et des chercheurs. Pour comparer les adresses, notre choix se porte sur les institutions (universités, grandes écoles) et non sur les laboratoires de recherche, d'abord parce que le nombre de laboratoires est très important comparé au nombre d'institutions, ensuite parce que l'ensemble des classements académiques mondiaux repose sur les institutions de recherche¹⁵.

À partir de l'information sur les adresses extraite de la base EconLit, nous avons effectué un travail conséquent d'harmonisation des noms et, concernant la France, de regroupement de structures¹⁶. Pour les chercheurs appartenant à plusieurs institutions, le contenu du champ adresse est séparé entre les différentes institutions. Ainsi une chercheuse qui signerait « Institution X et Université Y » obtiendrait deux champs adresse : champ 1 « Institution X » et champ 2 « Université Y ». Chaque structure recevra ensuite un score qui dépendra de la métrique choisie.

Deux métriques de classement différentes sont dès lors mises en œuvre. La première consiste à ajouter une unité à une institution pour chacune de ses apparitions (encadré 1). Bien qu'un article puisse être l'œuvre de plusieurs chercheurs et que chacun puisse signer de plusieurs institutions, chaque apparition rapportera ainsi un point à l'institution concernée. Cette métrique s'attache surtout à la visibilité d'une institution, c'est-à-dire à sa capacité à s'associer à de nombreux travaux d'évaluation.

La seconde métrique repose sur la pondération des apparitions d'une institution par le nombre d'auteurs de l'article, ainsi que par le nombre d'institutions différentes dans l'adresse d'un auteur. Cette métrique assure que chaque article « vaut » au total 1 point (encadré 1). Puisque ce point est ensuite partagé entre auteurs et entre institutions, elle favorise les articles avec peu d'auteurs et les chercheurs rattachés à une seule institution¹⁷.

¹⁵ Par exemple, pour la France, le LIEPP est intégré à Sciences Po, et le GREQAM à l'université Aix-Marseille. Cela explique que nous ne repérons pas les laboratoires d'excellence (Labex) sur l'évaluation des politiques publiques financés dans le cadre des programmes d'investissement d'avenir (PIA).

¹⁶ Par exemple, nous avons fusionné dans une seule institution le CREST, l'ENSAE et l'INSEE. De la même manière, nous avons regroupé tous les sites de l'INRA en une seule institution. En Allemagne, nous avons regroupé tous les instituts Max Planck sous une seule bannière.

¹⁷ Le calcul de cette métrique implique que tous les champs « auteur » et « adresse » d'un article soient correctement renseignés, et en particulier que le nombre d'auteurs soit égal au nombre d'adresses répertoriées. Cette restriction amène à exclure du calcul environ 10 % des articles qui ne vérifient pas cette condition. Les résultats comparés des deux métriques montrent cependant que cette condition n'affecte pas nos conclusions.

Encadré 1 – Illustration des deux métriques utilisées

Considérons un article signé par trois auteurs :

Duhautois, Richard; Emmanuelle Walkowiak et Oana Calavrezo.(2009),. "The Substitution of Worksharing and Short-Time Compensation in France: A Difference-in-Differences Approach", Economics Bulletin, 29(2), 820-33.

Il y a alors trois champs adresse dans la base EconLit pour chacun des trois auteurs :

- U Paris XII, Est-Marne-la-Vallee and CREST ;
- LEO, U Orleans and CEE ;
- U Paris XII, Est-Marne-la-Vallee.

Selon la première métrique (métrique simple), Paris XII obtient deux points, tandis que le CREST-INSEE-ENSAE, l'université d'Orléans et le CEE reçoivent chacun un point.

Selon la seconde métrique (pondérée), chacune des institutions rattachées à l'article reçoit le poids suivant :

- Paris XII : $\frac{1}{3} \times \frac{1}{2} + \frac{1}{3} = \frac{1}{2}$
- CREST-INSEE-ENSAE : $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$
- Université d'Orléans : $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$
- CEE : $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$

Nous calculons ainsi le classement des cinq institutions qui produisent le plus d'évaluations d'impact par pays selon ces deux métriques distinctes. La Figure 15 présente le classement des institutions à partir de la métrique simple et la Figure 16 à partir de la métrique pondérée (voir l'annexe E pour le top 20 des institutions, avec ou sans pondération). Il est important de remarquer que les résultats avec la seconde métrique, pourtant plus complexe à mettre en œuvre, sont extrêmement semblables à ce que l'on obtient sans pondération. Nos classements sont donc robustes au changement de métrique et nous interprétons les résultats dans leur ensemble.

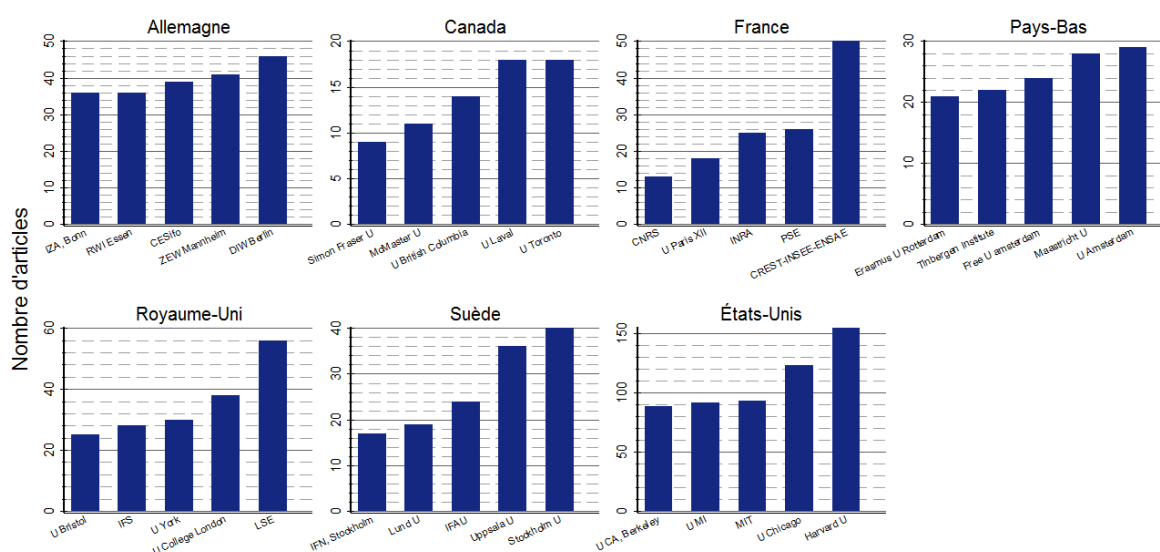
En premier lieu, nous constatons qu'aucun pays de notre échantillon ne présente d'institution étrangère dans le top 5. Cette observation frappante confirme le constat précédent sur le caractère national de l'évaluation : dans tous les pays, une large part des études sont produites par des chercheurs affiliés à des institutions de recherche nationales. En outre, les institutions qui dominent le classement dans tous les pays de l'échantillon sont avant tout les acteurs académiques nationaux les plus reconnus : on peut par exemple citer l'École d'économie de Paris pour la France ; l'université d'Amsterdam aux Pays-Bas ; la London School of Economics et l'University College London au Royaume-Uni ; les universités d'Harvard, de Chicago et de Pennsylvanie ainsi que le Massachusetts Institute of Technology pour les États-Unis. À l'inverse, nous ne voyons pas apparaître dans ces classements d'acteur spécialisé dans le domaine de l'évaluation. Les évaluations d'impact sont donc largement un produit des grandes institutions de recherche en économie.

Dans ce paysage, la France se démarque néanmoins du reste de l'échantillon, dominé par la présence des universités. Le paysage français de l'évaluation fait en effet intervenir d'autres acteurs plus proches du milieu administratif ; ainsi, l'ensemble formé par le CREST-INSEE-ENSAE apparaît systématiquement au premier rang de notre classement. Ce constat s'explique probablement par la spécificité du service statistique public français, qui à la fois produit des données et réalise des études. Par ailleurs, l'INSEE et les services statistiques ministériels ont longtemps bénéficié d'un accès quasi exclusif aux données individuelles, avant la mise en place du centre d'accès sécurisé à distance (CASD) qui a permis aux chercheurs et aux étudiants d'exploiter à leur tour ces données sans avoir à travailler sur place dans les institutions productrices. En outre, les grandes institutions nationales de recherche telles que le CNRS¹⁸ et l'INRA se placent également parmi les plus grands acteurs de l'évaluation en France.

Dans plusieurs pays de l'échantillon, des organismes de recherche indépendants sont également très présents dans le paysage de l'évaluation. C'est par exemple le cas de l'Institute for Fiscal Studies (IFS) au Royaume-Uni, institut de recherche spécialisé sur les questions fiscales et budgétaires mais qui réalise des évaluations sur des domaines variés (éducation, marché du travail, santé, retraites, inégalités...). On peut aussi citer à ce titre le Research Institute of Industrial Economics (IFN) en Suède, historiquement spécialisé dans l'étude des politiques industrielles mais qui produit aujourd'hui des travaux d'évaluation d'impact et d'économie expérimentale, ou encore le Tinbergen Institute aux Pays-Bas reconnu pour ses programmes d'enseignement et de recherche en économie, économétrie et finance. Leur présence dans ce classement témoigne de l'existence, dans ces pays, d'acteurs indépendants manifestant un grand intérêt pour l'évaluation des politiques publiques et également capables de mobiliser des ressources et des méthodes scientifiques pour effectuer ces analyses. L'Allemagne connaît à cet égard une situation particulièrement atypique puisque la quasi-totalité des acteurs de son top 5 sont des instituts de recherche indépendants des universités mais qui utilisent le savoir-faire des chercheurs d'Allemagne ou de l'étranger. Ces organismes peuvent être généralistes (comme le DIW, le ZEW ou le RWI) ou spécialisés (par exemple l'IZA et l'IAB sur les questions relatives au marché du travail) et sont régulièrement sollicités par les ministères afin de réaliser des évaluations de politiques publiques.

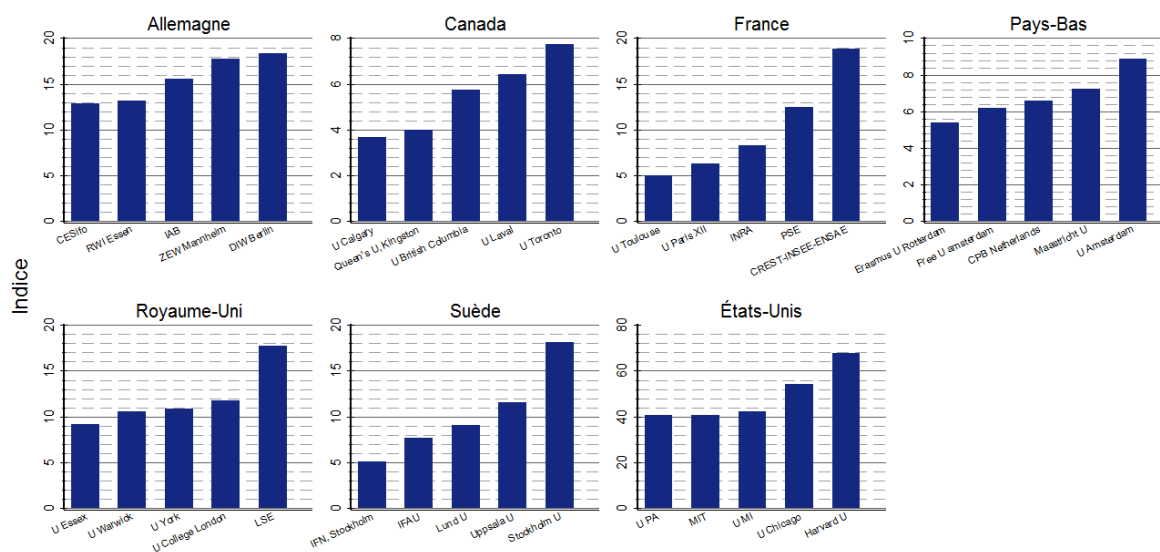
¹⁸ Dans le cas où plusieurs affiliations sont indiquées, parmi lesquelles le CNRS, la première affiliation du chercheur est favorisée. Par exemple, un chercheur dont l'affiliation indiquée serait « CNRS (DELTA) » est alors rattaché au CNRS, tandis qu'un chercheur affilié « PSE-CNRS » est rattaché à PSE.

Figure 15 – Top 5 des institutions par pays, métrique simple



Note de lecture : Pour les États-Unis, l'université d'État du Michigan est désignée par U MI et l'université de Californie à Berkeley par U CA, Berkeley.

Figure 16 – Top 5 des institutions par pays, métrique pondérée par auteur et nombre d'institutions



Note de lecture : Pour les États-Unis, l'université d'État du Michigan est désignée par U MI et l'université de Pennsylvanie par U PA.

L'analyse plus détaillée de la distribution des institutions dans la production d'évaluations d'impact (Tableau 2 pour la métrique sans pondération et Tableau 3 pour la métrique avec pondération) fait apparaître une certaine dispersion de la production : plus de la moitié des institutions n'apparaissent qu'une seule fois parmi les études portant sur chaque pays. Seule

une minorité d'institutions intervient donc de manière répétée sur le champ de l'évaluation d'impact. Les 5 % d'institutions les plus représentées dans les pays affichent le plus souvent plus de 12 occurrences, ou 4 à 5 occurrences pondérées ; ce seuil est cependant beaucoup plus élevé pour les États-Unis, où les institutions appartenant au 95^e centile ont au minimum 18 occurrences ou 8,7 occurrences pondérées.

La comparaison de la part des tops 10, 5 et 1 d'institutions dans la production d'évaluations d'impact entre les pays fait apparaître des variations importantes dans la concentration de cette production. Les États-Unis sont le pays où la production est la moins concentrée sur les premières institutions en raison de l'important volume total de production : la part du top 5 d'institutions ne représente qu'environ 10 % du total. Inversement, la Suède et les Pays-Bas, qui ont les productions les plus modestes de notre échantillon, voient plus de 40 % de leur production concentrée autour des cinq premières institutions. Les pays du reste de l'échantillon (France, Allemagne, Canada, Royaume-Uni) se situent entre ces deux extrêmes avec une part du top 5 comprise entre 20 % et 30 %. Le calcul d'une mesure synthétique de la concentration de la production au niveau des institutions (indice de Herfindahl-Hirschmann), confirme ces constats.

En termes de concentration, la France se place donc dans le groupe médian de notre échantillon. Cependant, au sein de ce groupe, elle présente la concentration la plus importante : la part d'évaluations produite par le top 10 dépasse les 40 %, tandis qu'elle avoisine plutôt 35 % dans les trois autres pays. En outre, le premier producteur d'évaluations (le CREST-INSEE-ENSAE) représente plus de 10 % du total, comme en Suède et aux Pays-Bas, tandis que dans les autres pays cette part ne dépasse pas 7 %.

Tableau 2 – Distribution des publications dans les institutions non pondérées

	Statistiques descriptives		Centile			Part des top institutions			Concentration
	Nombre d'institutions	Moyenne	50%	90%	95%	Top 10	Top 5	Top 1	IHH
France	170	2,8	1	5	12	40,7	27,9	10,6	0,026
Allemagne	249	3,2	1	6	12	38,6	25,0	5,8	0,020
Pays-Bas	80	3,5	1	11	21,5	64,4	44,1	10,3	0,051
Suède	104	3,2	1	8	13	56,5	40,8	12,0	0,046
Canada	135	2,3	1	5	7	32,8	22,7	5,8	0,018
Royaume-Uni	270	2,9	1	6	12	33,6	22,5	7,1	0,017
États-Unis	1204	4,5	1	10	18	17,1	10,1	2,8	0,006

Note de lecture : Dans la première colonne est reporté le nombre d'institutions différentes qui ont publié dans le pays. La deuxième colonne présente la moyenne du nombre d'occurrences des institutions dans le pays. Par exemple pour la France, les 170 institutions différentes ayant produit les évaluations d'impact apparaissent en moyenne 2,8 fois chacune dans la base. Les trois colonnes suivantes présentent les médianes des distributions ainsi que les 90^e et 95^e centiles. Les 5 % des institutions françaises les plus présentes dans la base ont toutes plus de 12 occurrences par institution. Les trois colonnes suivantes reportent la part des occurrences issues des premières institutions : top 10, top 5 et top 1. Pour la France, les 10 premières institutions représentent 40,7 % des adresses, les 5 premières 27,9 % et la première 10,6 %. Enfin dans la dernière colonne nous calculons l'indice de Herfindahl-Hirschmann associé à la distribution (plus il est élevé, plus la distribution est concentrée).

Source : EconLit, traitement des auteurs

Tableau 3 – Distribution des publications dans les institutions pondérées

	Statistiques descriptives		Centile			Part des top institutions			Concentration
	Nombre d'institutions	Moyenne	50%	90%	95%	Top 10	Top 5	Top 1	IHH
France	170	1,0	0,3	1,8	3,8	43,7	30,7	11,7	0,032
Allemagne	249	1,3	0,5	2,5	5,3	38,3	22,4	5,8	0,020
Pays-Bas	80	1,0	0,5	2,8	5,8	62,4	39,6	11,1	0,050
Suède	104	1,1	0,3	2,2	5,1	57,9	45,2	15,9	0,056
Canada	135	1,0	0,5	2,4	3,5	32,3	20,2	5,0	0,019
Royaume-Uni	270	1,2	0,5	2,9	4,5	29,2	17,3	5,5	0,015
États-Unis	1204	2,1	0,5	4,2	8,7	16,4	9,8	2,7	0,006

Note de lecture : La lecture est la même que le tableau précédent, mais la métrique utilisée est la métrique pondérée. Par conséquent la distribution considérée n'est pas celle des occurrences des institutions dans l'ensemble des adresses mais celles des occurrences pondérées par nombre d'auteurs et d'institutions, soit le nombre d'articles associés à chaque institution. Selon cette métrique, les 170 institutions ayant produit des études sur la France ont en moyenne produit l'équivalent d'un article chacune. Les 10 premières institutions représentent 43,7 % des articles, les 5 premières 30,7 % et la première 11,7 %.

Source : EconLit, traitement des auteurs

Conclusion

Les évaluations d'impact se sont largement développées dans tous les pays étudiés à partir des années 2000 (et bien avant pour les États-Unis avec l'évaluation dans les années 1960 des programmes sociaux de lutte contre la pauvreté du président Johnson). La France présente une production relativement modeste, mais connaît une accélération dans la période récente, ce qui lui permet de rattraper une partie de son retard vis-à-vis d'autres pays plus avancés comme l'Allemagne ou le Royaume-Uni. Les évaluations dans les pays européens couvrent majoritairement les domaines de l'emploi, de l'éducation et de la santé. La France, comme l'Allemagne et la Suède, s'intéresse en priorité aux questions d'emploi, à la différence des pays anglo-saxons, davantage spécialisés sur le domaine de la santé. Les producteurs majeurs d'évaluations d'impact sont bien souvent les principaux acteurs académiques nationaux ou les grands organismes de recherche indépendants qui emploient des chercheurs ou des universitaires ; la France se démarque cependant par la présence importante d'acteurs à mi-chemin de l'administration et de la recherche publique. Elle publie également plus souvent que les autres pays les évaluations d'impact dans les revues nationales, ce qui peut en limiter la visibilité internationale, mais à l'inverse en accroître la réception nationale. On peut toutefois mentionner le cas de la revue *Économie et Statistique* qui, depuis 2017, publie tous ses articles en français et en anglais dans l'objectif d'accroître la diffusion des articles français à l'étranger.

Cette étude dresse un certain nombre de constats sur la place de la France dans la production d'évaluations d'impact, sans toutefois les expliquer. Plusieurs hypothèses peuvent être avancées pour expliquer le classement relativement défavorable de la France, notamment par rapport aux pays d'Europe du Nord, au Royaume-Uni et à l'Allemagne : une moindre culture de l'évaluation des politiques publiques et une insuffisante formation des élites, notamment administratives et politiques, aux méthodes d'évaluation d'impact des politiques publiques ; un défaut d'impulsion politique et un rôle faible du Parlement dans l'évaluation des politiques publiques en France ; un accès des chercheurs aux données individuelles plus tardif et plus complexe (Bozio, Geoffard, 2017) ; une offre d'évaluateurs plus restreinte, sans doute en lien avec les facteurs précédents (moindre demande d'évaluations, accès plus tardif aux micro-données).

Cette première étude s'inscrit dans un projet plus global visant à classer les sept pays étudiés (Allemagne, Canada, États-Unis, France, Pays-Bas, Royaume-Uni et Suède) selon deux dimensions : leur capacité à produire des évaluations d'impact et leur capacité à les utiliser. La deuxième dimension sera l'objet d'une autre étude de nature qualitative qui analysera la place des évaluations d'impact et des chercheurs qui les produisent dans les processus de décision.

Références

- Bachtler J., Polverari L., Taylor S., Ashcroft B. et Swales K. (2000), *Methodologies used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment*, Final Report to the Scottish Executive, European Policies Research Centre and Fraser of Allander Institute, septembre.
- Bénassy-Quéré A., Blanchard O. J. et Tirole J. (2017), « Les économistes dans la cité », *Les notes du conseil d'analyse économique*, n° 42, juillet.
- Bosquet C. et Combes P.-P. (2012), « Un panorama de la recherche française en économie comparant les approches Google Scholar et EconLit », *Revue d'économie politique*, 2012/4, Vol. 122, p. 477-545, octobre.
- Bozio A. et Geoffard P.-Y. (2017), *L'accès des chercheurs aux données administratives. État des lieux et propositions d'actions*, rapport au secrétaire d'État chargé de l'Industrie, du Numérique et de l'Innovation, groupe de travail du CNIS, mars.
- Bozio A. et Romanello L. (2017), « Évaluation des politiques publiques : le bilan contrasté du quinquennat », *Les notes de l'IPP*, n° 25, mars.
- Combes P.-P., Linnemer L. (2001), « La publication d'articles de recherche en économie en France », *Annales d'économie et de statistique*, n° 62, p. 5-47, avril-juin.
- Combes P.-P. et Linnemer L. (2003), « Where are the economists who publish? Publication concentration and rankings in Europe based on cumulative publications », *Journal of the European Economic Association*, 1(6), p. 1250-1308, décembre.
- Desplat R. et Ferracci M. (2016), *Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et des praticiens*, France Stratégie, septembre.
- Devaux-Spatarakis A. (2015), « The difficult encounter between experimental evaluation processes and stakeholders' interests », *Evaluation Connections*, décembre.
- Ellison G. (2002), « The slowdown of the economics publishing process », *Journal of Political Economy*, 110(5), p. 947-993.
- Esping-Andersen G. (1999), *Les trois mondes de l'État-providence*, Paris, PUF, p. 74, septembre.
- Evaluation Gap Working Group (2006), *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Washington DC, Center for Global Development, mai.
- Evidence Works (2016), *A Global Forum for Government*, rapport final de la conférence des 29 et 30 septembre 2016, décembre.
- Furubo J.-E., Rist R. C. et Sandahl R. (2002), *International Atlas of Evaluation*, New Brunswick & Londres, Transaction Publishers.
- Laurent C. et al., (2009), « Pourquoi s'intéresser à la notion d'"evidence-based policy" ? », *Revue Tiers Monde*, 2009/4, n° 200, p. 853-873, octobre-décembre.
- Lázaro B. (2015), « Comparative study on the institutionalisation of evaluation in Europe and Latin America », *Study n° 15, EUROsociAL*, juillet.
- Lubrano M., Bauwens L., Kirman A. et Protopopescu C. (2003), « Ranking economics departments in Europe: a statistical approach », *Journal of the European Economic Association*, 1(6), p. 1367-1401, décembre.

- Mirucki J. (2000), « Economics research in France: tentative conclusions based on EconLit Database », *Higher Education Management*, Vol. 12, n° 1, p. 85-103, mars.
- Nutley S. M., Davies H. T. O. et Tilley N. (2000), « Editorial: Getting research into practice », *Public Money & Management*, 20(4), p. 3-6, octobre.
- Nutley S. M. et Davies H. T. O. (2001), « Evidence-based policy and practice: moving from rhetoric to reality », *Third International, Inter-disciplinary Evidence-Based Policies and Indicator Systems Conference*, juillet.
- OECD (2015), *OECD Regulatory Policy Outlook 2015*, Paris, OECD Publishing, octobre.
- Polverari L. et Bachtler J. (2004), *Assessing the Evidence: the Evaluation of Regional Policy in Europe*, European Policy Research Paper, n° 56, European Policies Research Centre.
- Sanderson I. (2002), « Evaluation, policy learning and evidence-based policy making », *Public Administration*, 80(1), p. 1-22, décembre.
- Sharma M. R., Freudl D., Münch C., Schindler E. et Wegrich K. (2013), « Expert report on the implementation of ex-post evaluation: Good practice and experience in other countries », Prognos, Berlin.
- Sustainable Governance Indicators (2017), étude en ligne à l'adresse suivante : http://www.sgi-network.org/2017/Governance/Executive_Capacity/Evidence-based_Instruments
- Simha J. (2017), « Le laboratoire des politiques publiques. Réflexions sur la Garantie jeunes », *La vie des idées*, janvier.
- Varone F. et Jacob S. (2004), « Institutionnalisation de l'évaluation et nouvelle gestion publique : un état des lieux comparatif », *Revue internationale de politique comparée*, 11(2), p.272-292.

Annexes

A. Codes JEL de catégorie 1

Figure 17 – Code JEL de catégorie 1 les plus représentés

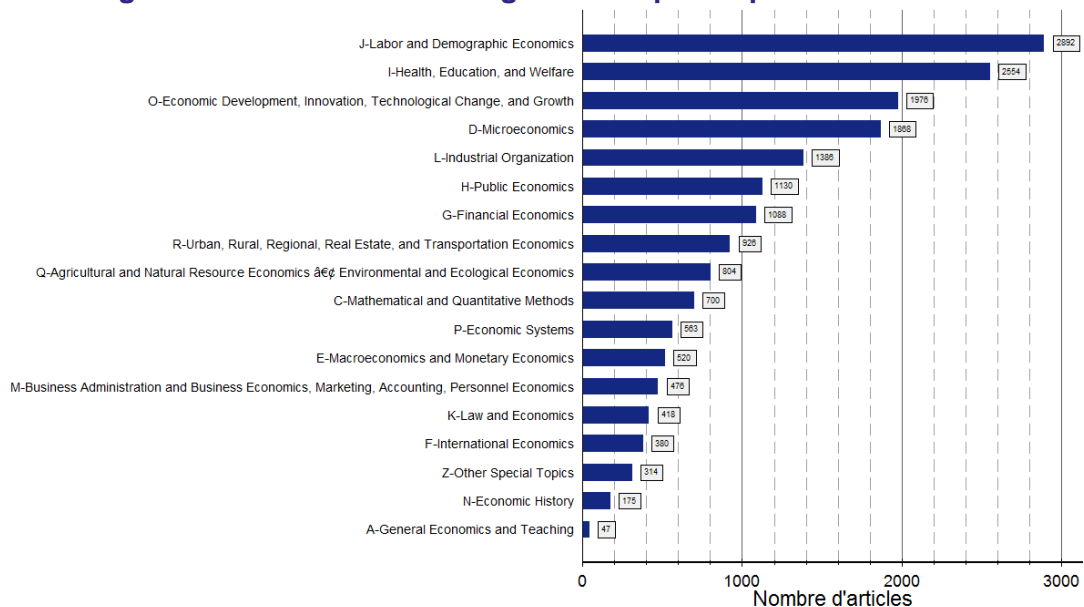
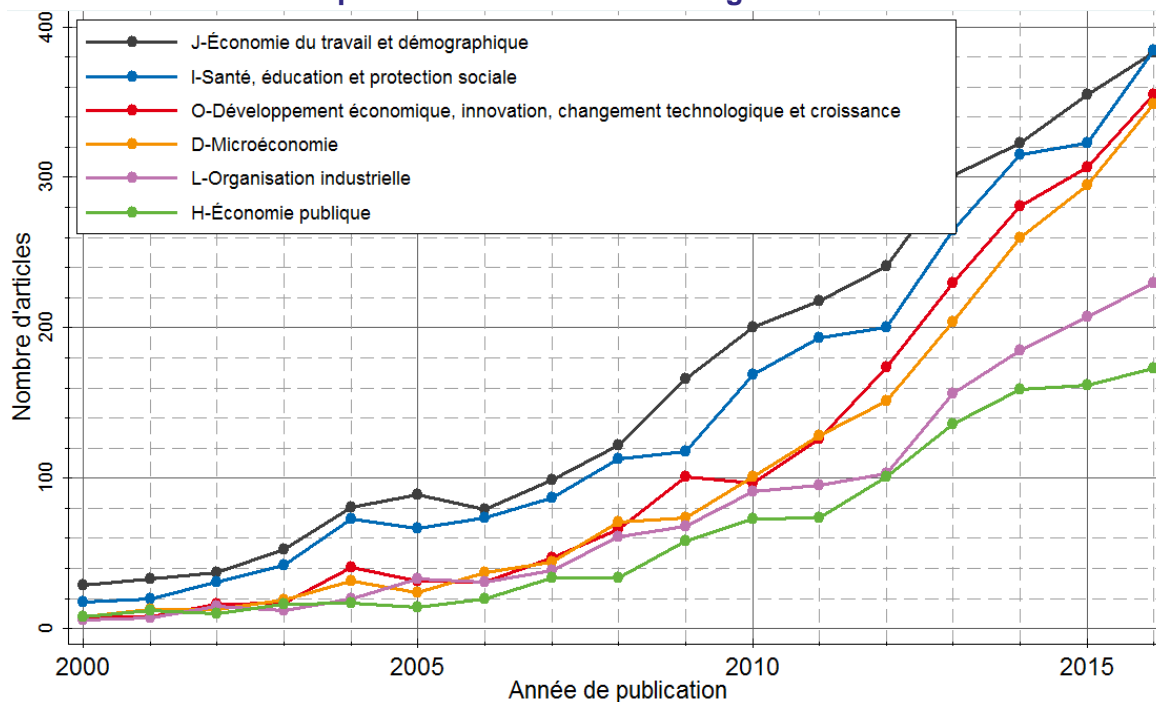
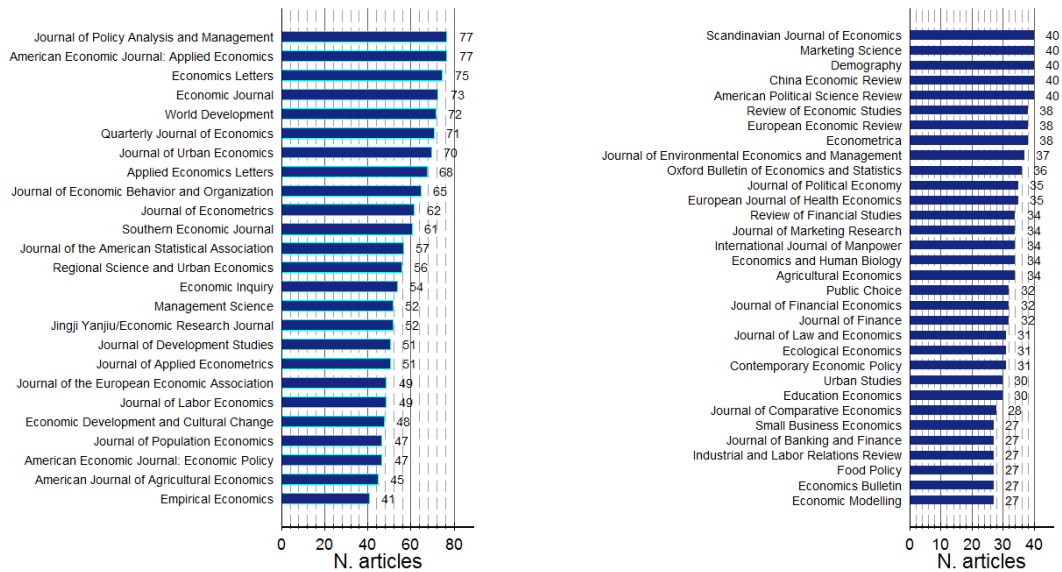


Figure 18 – Évolution des 6 premiers codes JEL de catégorie 1 entre 2000 et 2016



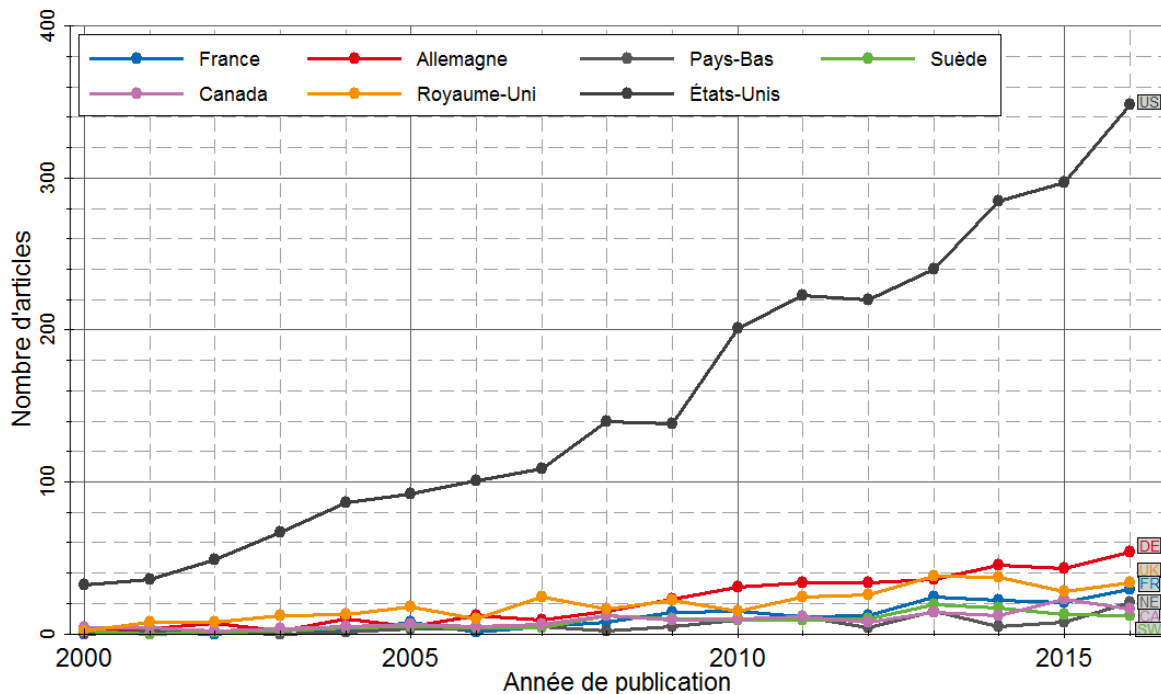
B. Classement des journaux

Figure 19 – Liste du classement général au-dessus de la 10^e place



C. Graphiques avec les États-Unis

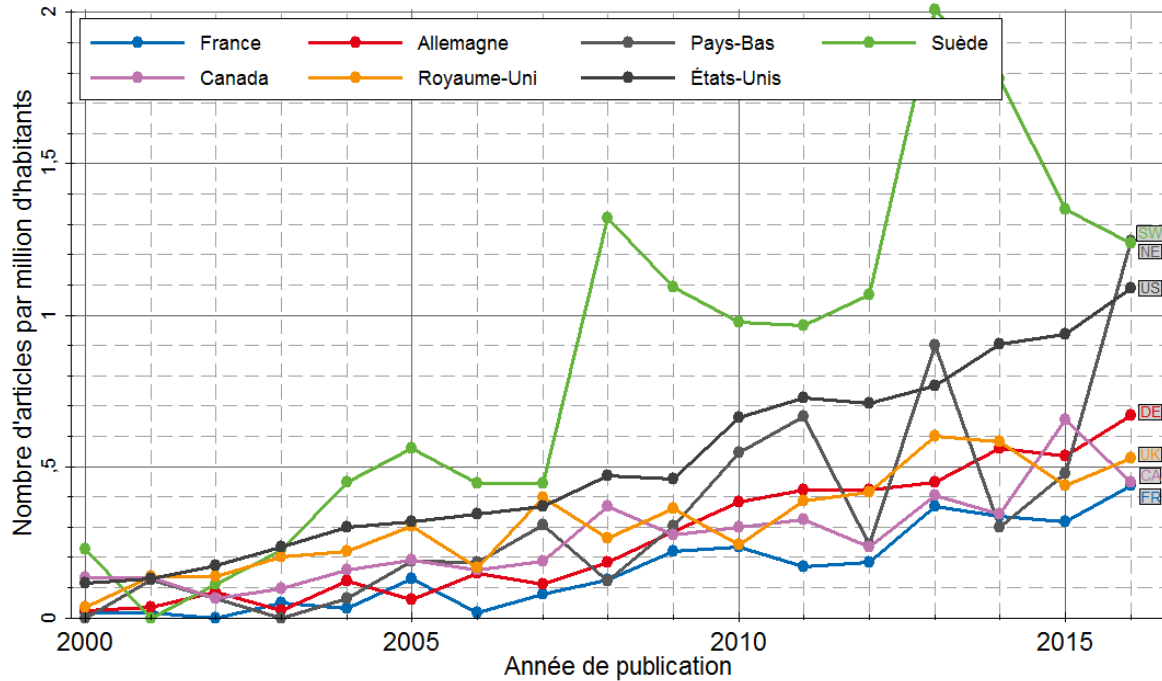
Figure 20 - Évolution du nombre d'articles publiés pour les pays de l'échantillon avec les États-Unis (2000-2016)



Source : EconLit, traitement des auteurs

D. Évolution rapportée à la population

Figure 21 – Évolution du nombre d'articles publiés rapporté à la population

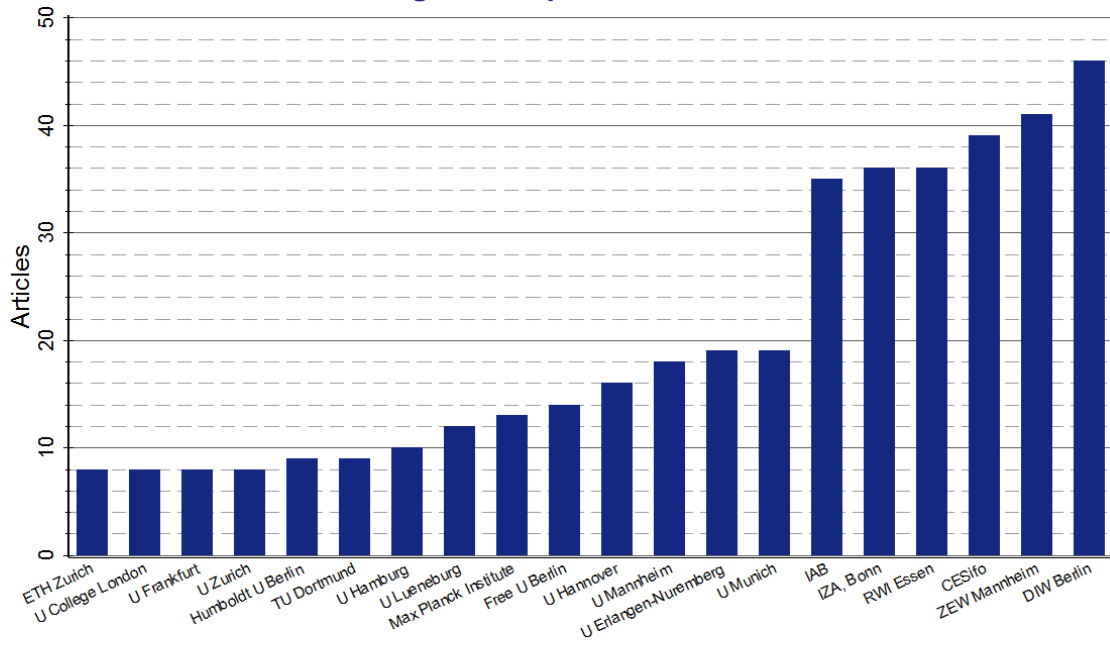


Note de lecture : Pour chaque année entre 2000 et 2016, on divise le nombre de publications par l'année $n - 2$. En effet, dans la version 9 de la Penn World Table, la population n'est présente que jusqu'en 2014.

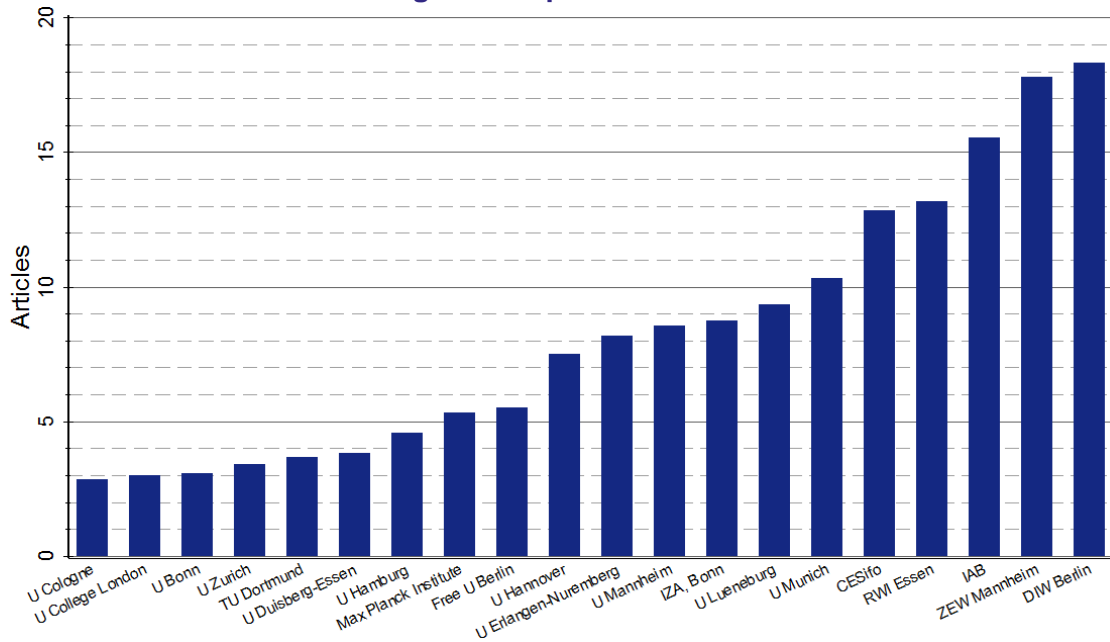
Source : EconLit, Penn World Table V9.0 pour la population en 2014, traitement des auteurs

E. Top 20 des institutions par pays

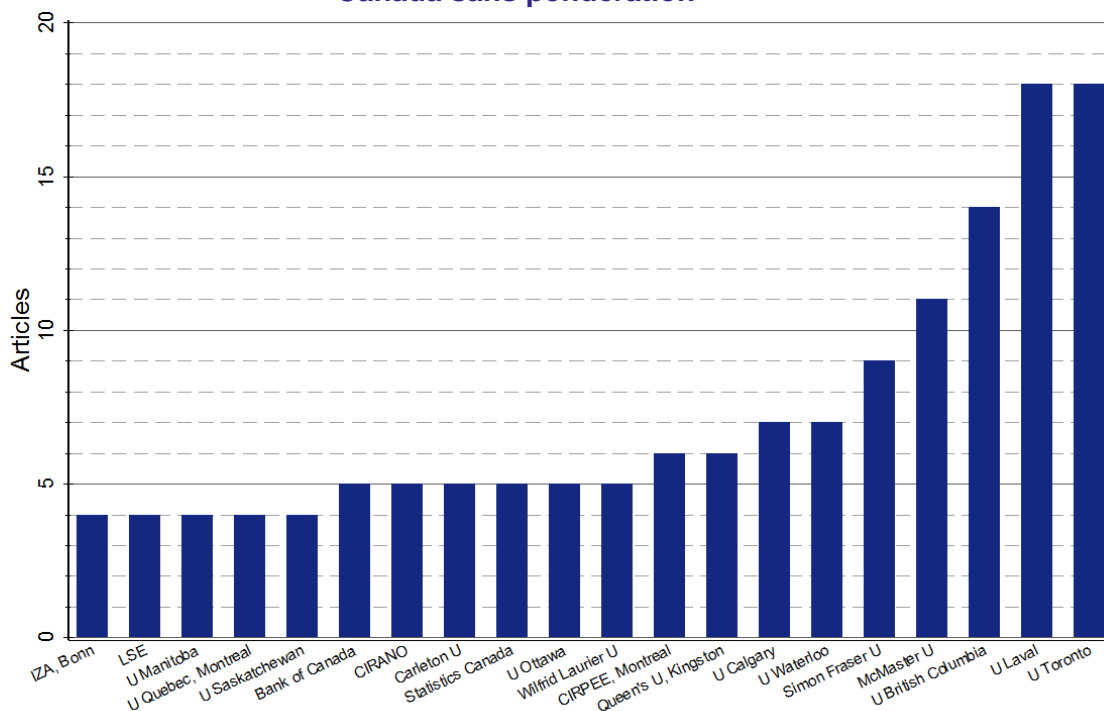
Allemagne sans pondération



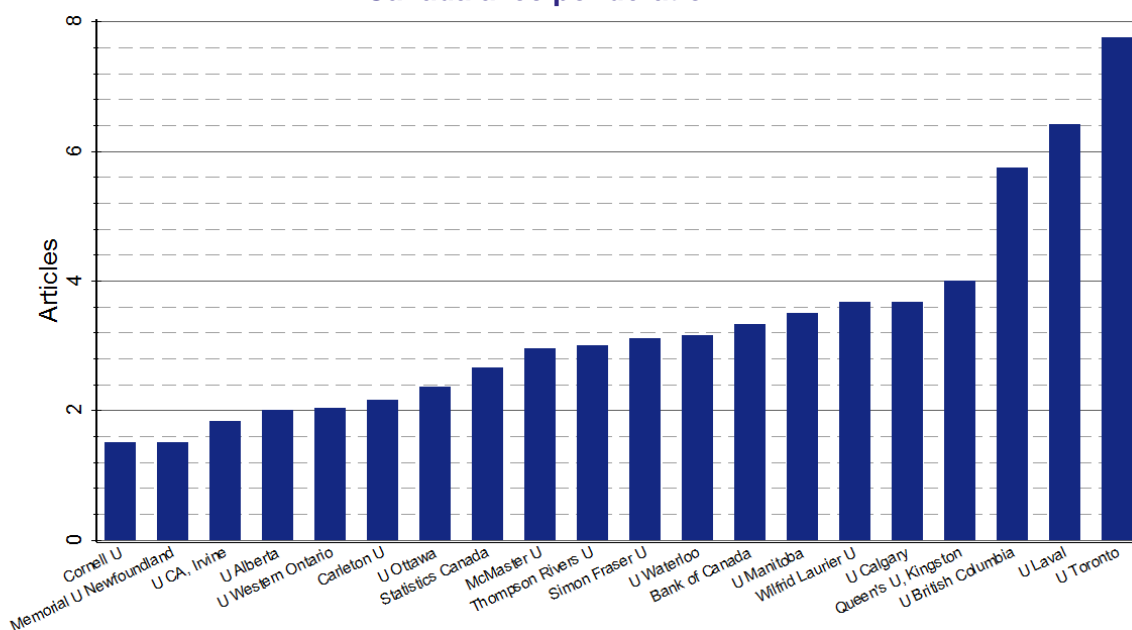
Allemagne avec pondération



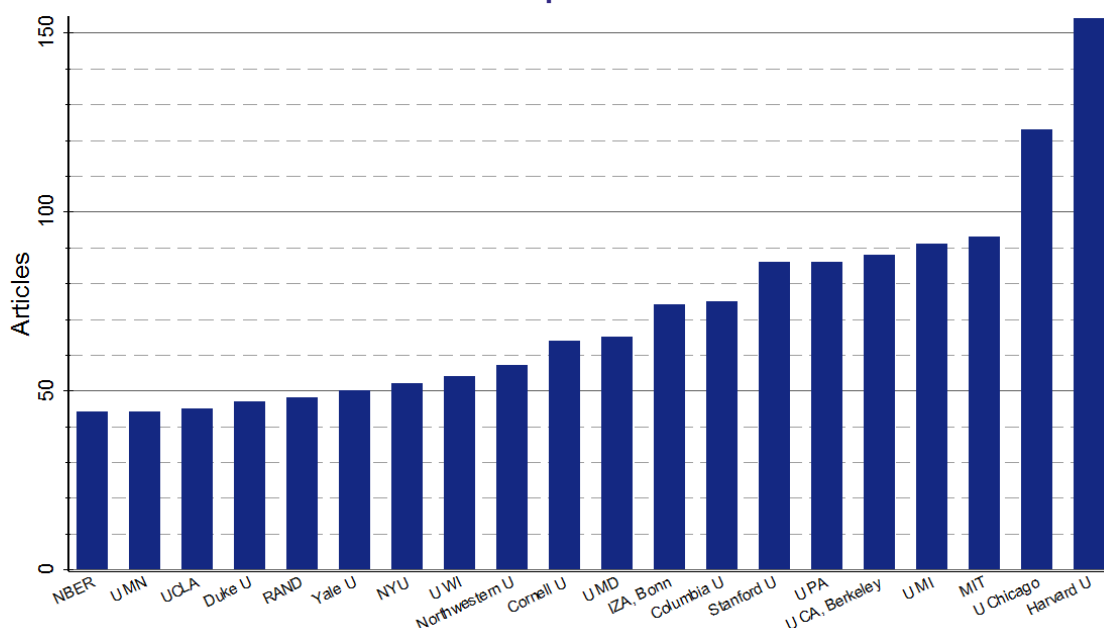
Canada sans pondération



Canada avec pondération

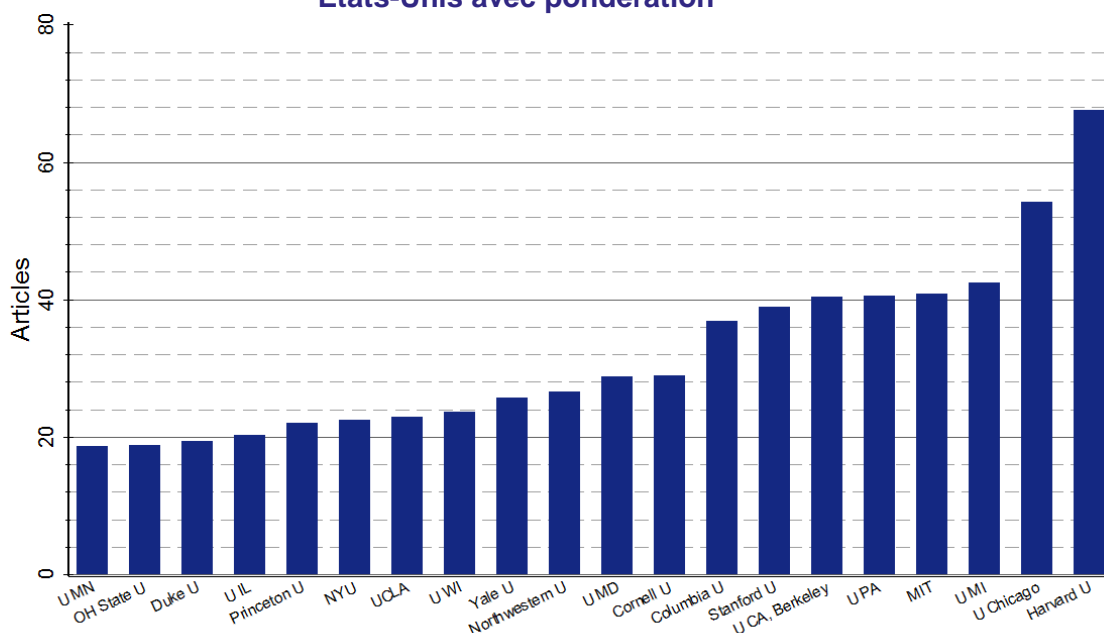


États-Unis sans pondération



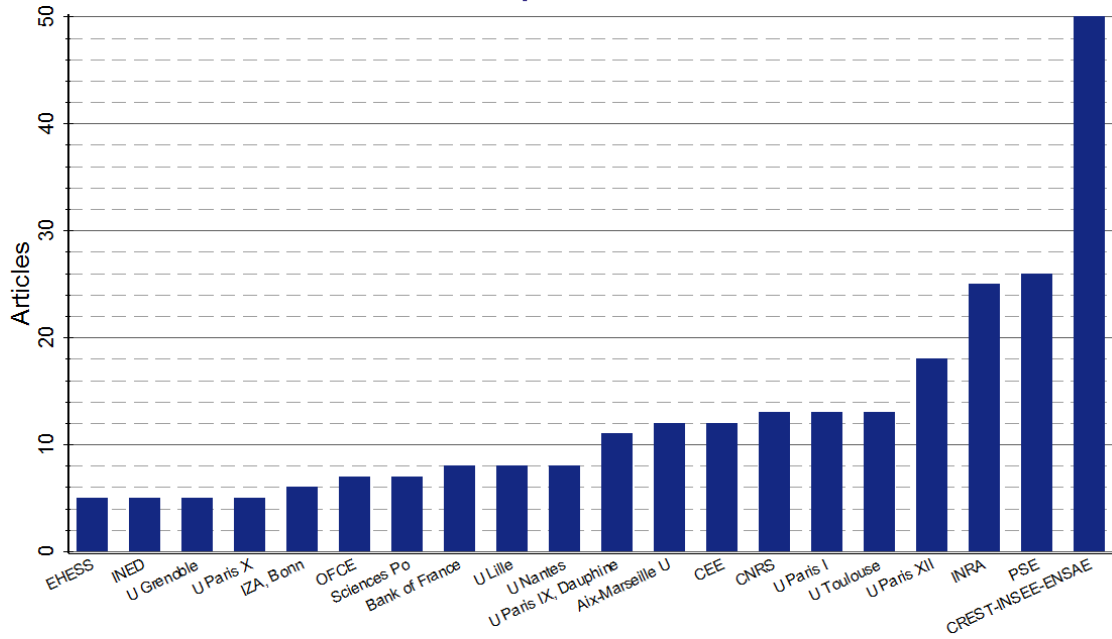
Note de lecture : L'université d'État du Michigan est désignée par U MI ; l'université de Californie à Berkeley par U CA, Berkeley ; l'université de Pennsylvanie par U PA ; l'université du Maryland par U MD ; l'université du Wisconsin à Madison par U WI ; l'université de New York par NYU ; l'université de Californie à Los Angeles par UCLA et l'université du Michigan par U MN.

États-Unis avec pondération

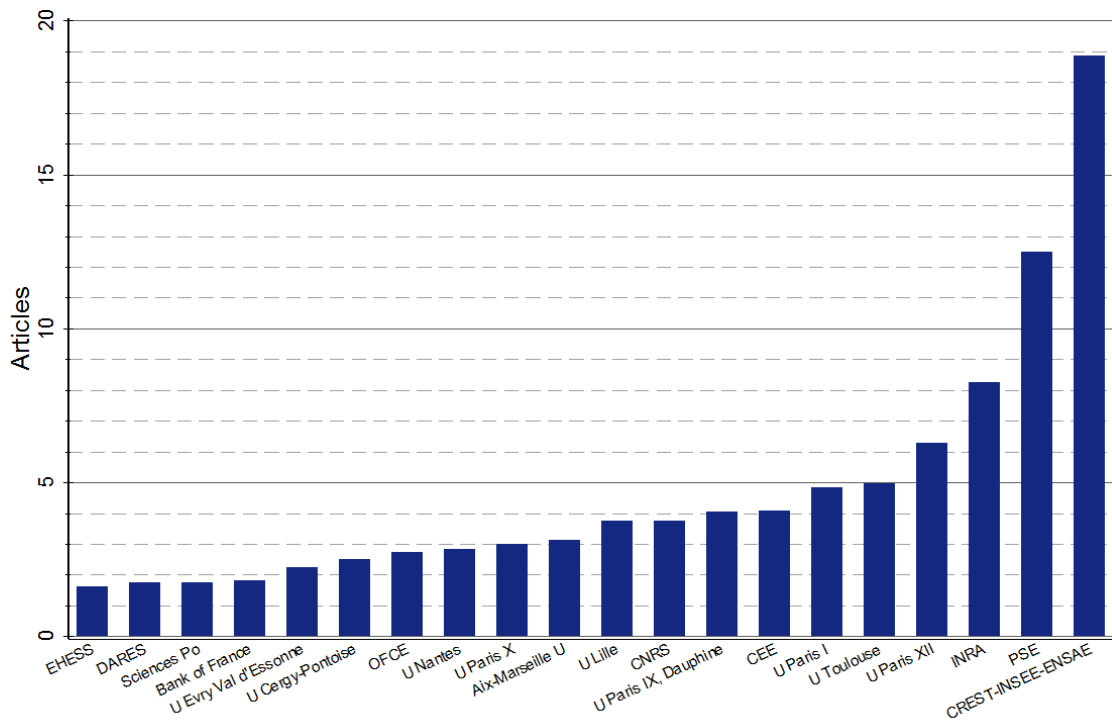


Note de lecture : L'université d'État du Michigan est désignée par U MI ; l'université de Pennsylvanie par U PA ; l'université de Californie à Berkeley par U CA, Berkeley ; l'université du Maryland par U MD ; l'université du Wisconsin à Madison par U WI ; l'université de Californie à Los Angeles par UCLA ; l'université de New York par NYU ; l'université d'Illinois par U IL ; l'université de l'Ohio par OH State U et l'université du Michigan par U MN.

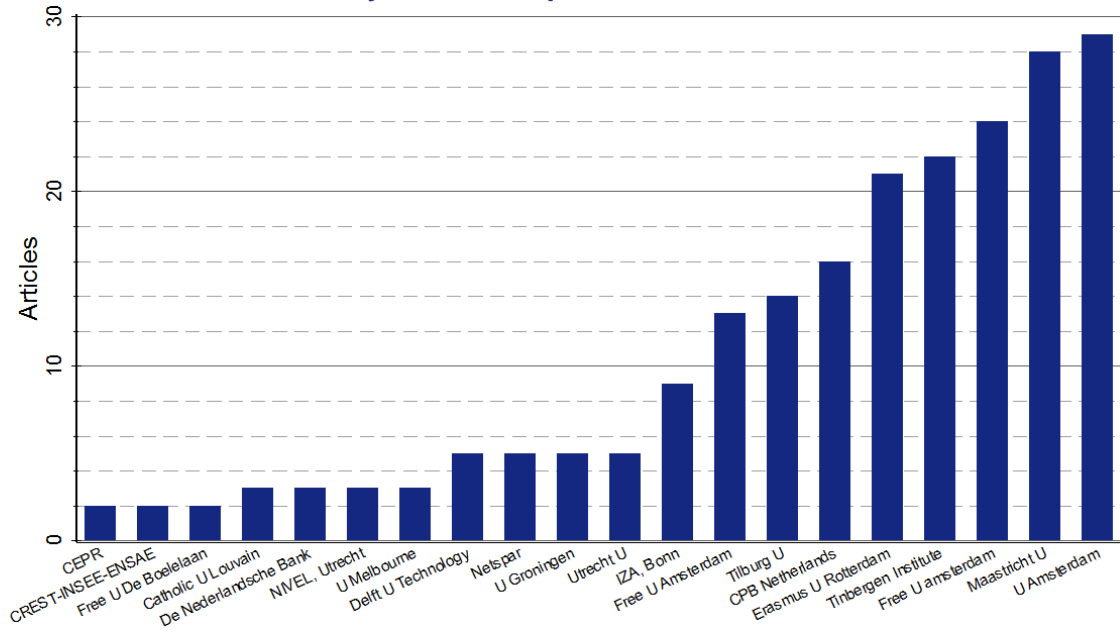
France sans pondération



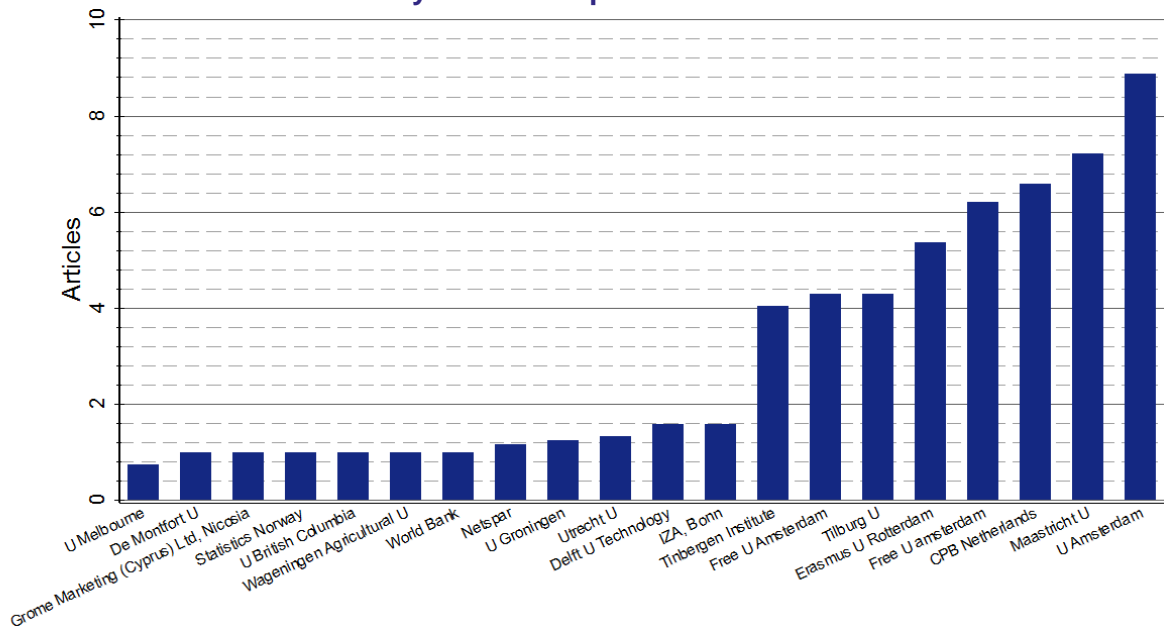
France avec pondération



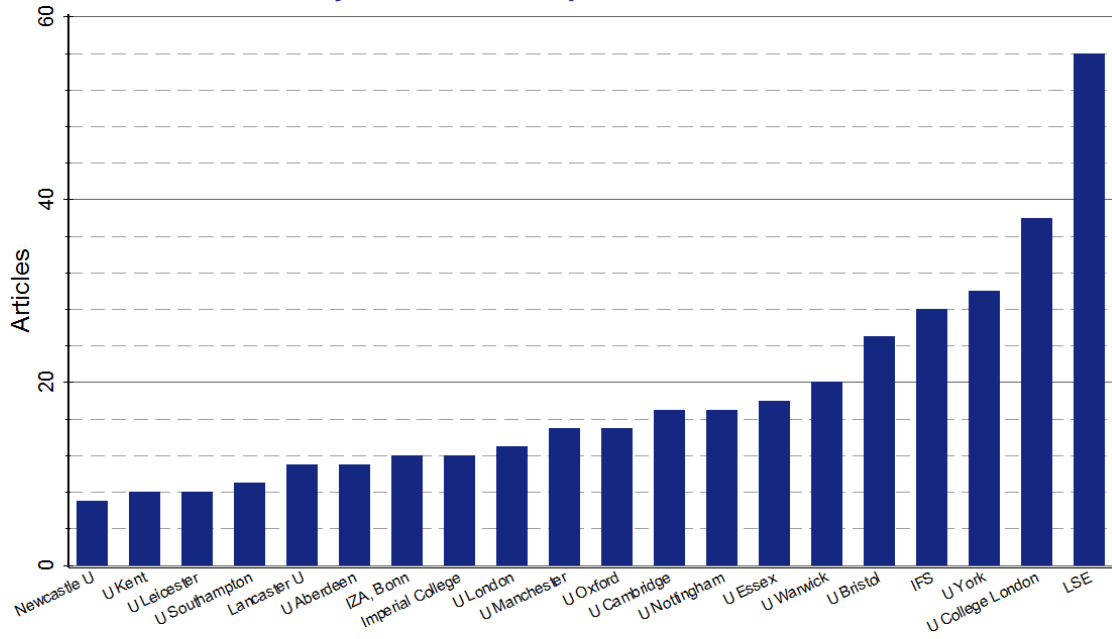
Pays-Bas sans pondération



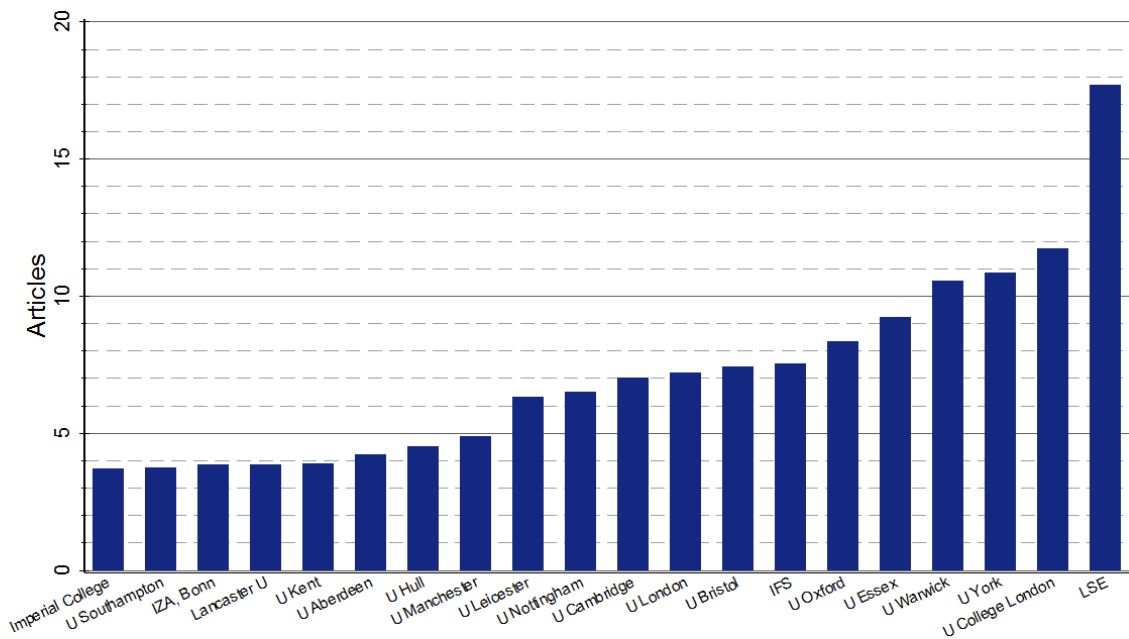
Pays-Bas avec pondération



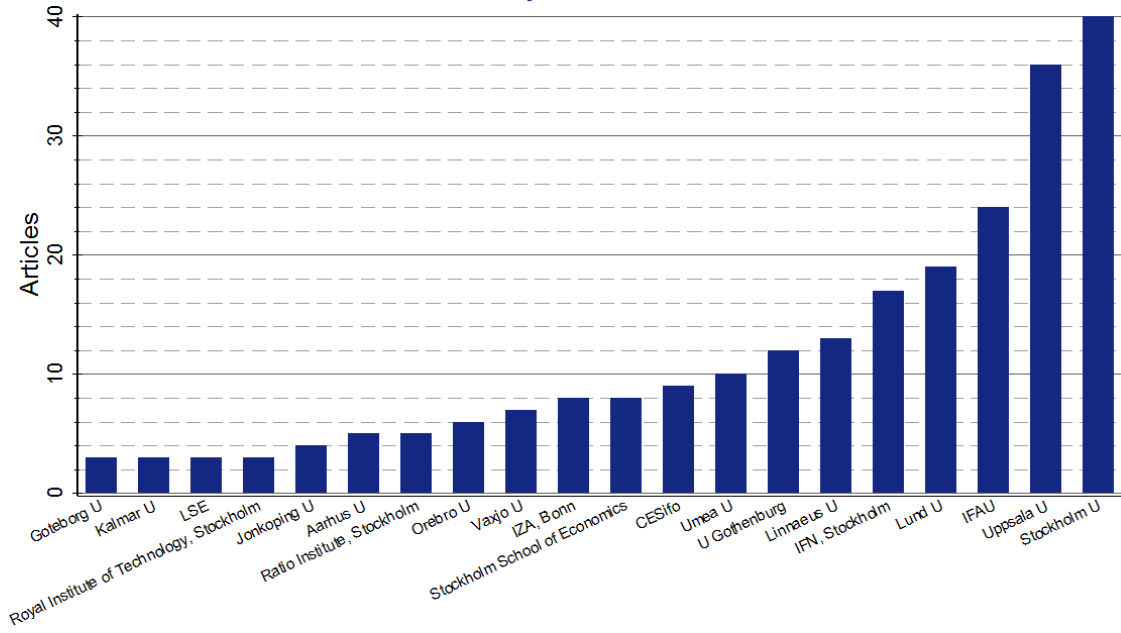
Royaume-Uni sans pondération



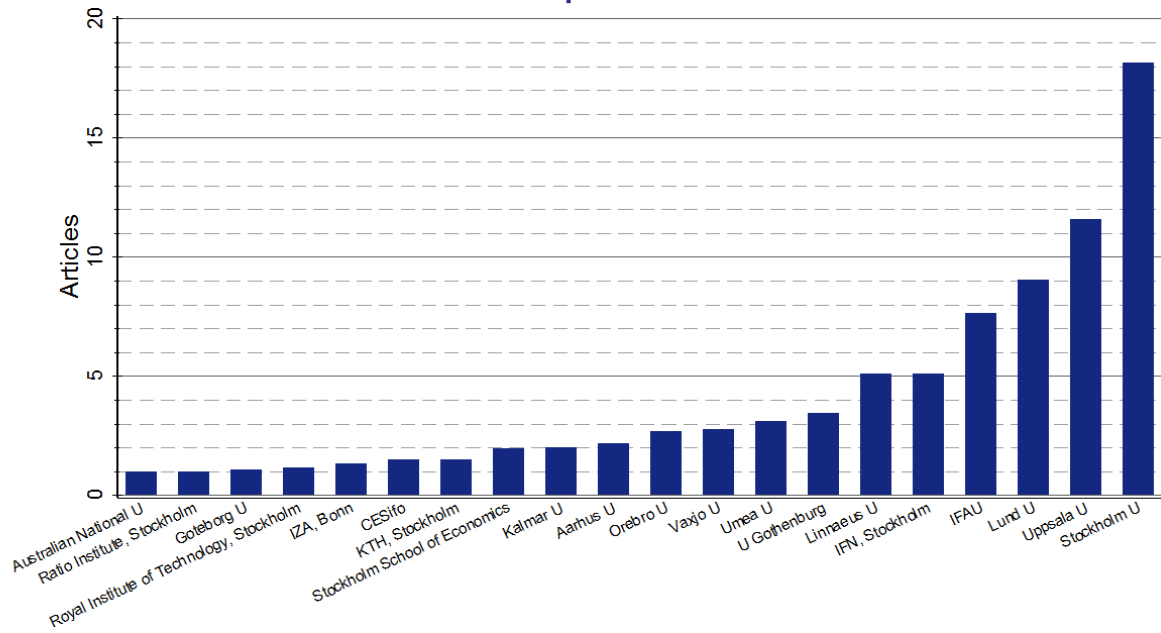
Royaume-Uni avec pondération



Suède sans pondération



Suède avec pondération



F. Tests sur la base d'articles relative à la France

Le champ de notre analyse est celui des **évaluations d'impact ex post réalisées à l'aide de méthodes quantitatives dites athéoriques**. Plus précisément, il s'agit d'études qui vérifient les critères suivants :

- l'objet de l'étude est une politique publique entendue au sens large (politique, programme, dispositif, etc.) ou un phénomène qui peut aider à la décision publique (exemples : la mobilité professionnelle, l'obtention d'un diplôme, l'obésité, la santé mentale, etc.) ;
- l'étude vise à quantifier les effets propres – causaux – de cette politique publique ou phénomène ;
- l'évaluation est réalisée *ex post* à l'aide de méthodes athéoriques sur données microéconomiques principalement.

Pour cibler ce champ, nous avons défini une requête à l'aide d'expressions portant sur les méthodes empiriques d'évaluation d'impact, à savoir :

- les expérimentations aléatoires contrôlées ;
- les quasi-expériences et expériences naturelles ;
- la régression sur discontinuité ;
- la méthode d'appariement ;
- les variables instrumentales ;
- les doubles ou triples différences ;
- les approches par régressions ;
- les effets causaux.

A contrario, sont exclus de notre ciblage les évaluations *ex ante* (dont la microsimulation), les évaluations qualitatives, les études descriptives (mise en évidence de corrélations), les articles en économie structurelle¹⁹ ou comportementale, et enfin les articles de synthèse ou purement méthodologiques.

Cette même requête est utilisée pour tous les pays de l'étude pour identifier les évaluations d'impact *ex post* dans la base EconLit. Malgré le soin apporté à la définition des termes de la requête, il peut exister une marge d'erreurs dans l'identification des articles de notre champ. Pour comprendre la nature des erreurs, nous avons mené deux types de tests pour la France uniquement. D'une part, nous avons examiné l'ensemble des articles identifiés pour la France (soit 180 articles) pour savoir dans quelle mesure ils comprennent des articles exclus de notre champ d'analyse (nous les appellerons alors des faux positifs). D'autre part, pour les revues françaises présentes dans la base EconLit, nous avons parcouru l'ensemble

¹⁹ Parmi les méthodes quantitatives, les méthodes structurelles peuvent également être utilisées pour évaluer l'impact *ex post* d'une politique publique. Cette catégorie de méthodes n'est pas ciblée dans notre analyse.

des articles publiés depuis l'entrée de ces revues dans la base pour examiner dans quelle mesure certains d'entre eux, bien qu'appartenant à notre champ, échappent à notre recherche (nous les appellerons les faux négatifs). Ce second test est beaucoup plus exigeant que le premier car il porte sur un nombre d'articles conséquent (5 300 articles examinés au total).

Ces erreurs, qu'elles jouent dans un sens ou dans l'autre (en surestimant ou en sous-estimant le nombre d'articles), peuvent également se produire pour les autres pays de l'étude.

1. Premier test pour la France : l'identification des articles dits faux positifs

Parmi les 180 articles de la base France, nous identifions 34 « faux positifs » qui sont précisés dans le tableau 1. Il s'agit principalement d'articles en économie structurelle et comportementale, d'articles de synthèse et de quelques articles descriptifs.

Tableau 1 : Articles « faux positifs » de la base France, classés par revue de publication (revues françaises puis internationales)

Année	Titre de l'article	Auteurs	Titre du journal
2001	Formation continue et carrières salariales : une évaluation sur données individuelles	Denis Fougère, Dominique Goux, Eric Maurin	<i>Annales d'économie et de statistique</i>
2013	Opportunities in higher education: an application to France	Brice Magdalou, Magali Jaoul-Grammare	
2016	'One Muslim is enough!' Evidence from a field experiment in France	Marie-Anne Valfort, David D. Laitin, Claire L. Adida	
2016	Spatial competition in the French supermarket industry	Stéphane Turolla	
2013	Évaluation des aménités urbaines par la méthode des prix hédoniques : une application au cas de la ville d'Angers	Muriel Travers, Solène Larue, Gildas Appere	<i>Économie et statistique</i>
2014	L'expérimentation « telle qu'elle se fait » : leçons de trois expérimentations par assignation aléatoire	Agathe Devaux-Spatarakis	<i>Formation emploi : revue française de sciences sociales</i>
2007	Individual responsiveness to information in CV surveys: commitment matters	Susan Cleary, Stéphane Luchini, Olivier Chanel	<i>Revue d'économie politique</i>
2011	Utilisation et déclasserement du capital	Guillaume Horny, Antoine Bonleu, Gilbert Cette	
2016	Effets de quartier, effet de département: discrimination liée au lieu de résidence et accès à l'emploi	Emilia Ene, Yannick L'Horty, Pascale Petit, Mathieu Bunel	<i>Revue économique</i>
2007	Modèles politico-économétriques et prévisions électorales pour mai 2007	François Facchini, Antoine Auberger, Jean-Dominique Lafay	<i>Revue française d'économie</i>
2009	Pollution diffuse aux nitrates et incitations économiques	Sandrine Spaeter, Alban Verchere, François Cochard, Marc Willinger	
2011	Évaluation aléatoire et expérimentations sociales	Pascale Petit, Yannick L'Horty	
2014	La discrimination à l'entrée des établissements scolaires privés	Pascale Petit, Thomas Brodaty, Loïc du Parquet	
2013	L'expérimentation sociale aléatoire en France en trois questions	Bernard Gomel, Evelyne Serverin	<i>Travail et emploi</i>
2015	Three major issues concerning randomised social experimentations in France	Bernard Gomel, Evelyne Serverin	
2013	L'expérimentation dans l'action publique. Usages, pratiques et jugement des acteurs	Jules Simha, Carole Tuchsirer, François Sarfati, Marie-Christine Bureau	

2010	Location intentions of immigrants at retirement: stay/return or go 'back and forth'?	Augustin de Coulon, François-Charles Wolff	<i>Applied Economics</i>
2014	Is financial repression a solution to reduce fiscal vulnerability? The example of France since the end of world war II	Anne Peguin-Feissolle, Gilles Dufrenot, Marcel Aloy	
2008	A dual-target monetary policy rule for open economies: an application to France	C. James Huen, Ruey Yau	<i>Applied Economics Letters</i>
2012	Generalised Taylor and generalised Calvo price and wage setting: micro-evidence with macro implications	Hervé Le Bilan, Huw Dixon	<i>Economic Journal</i>
2013	Reversal and magnitude effects in long-term time preferences: results from a field experiment	Florence Jacquet, Géraldine Bocqueho, Arnaud Reynaud	<i>Economics Letters</i>
2000	Private consumption behavior, liquidity constraints and financial deregulation in France: a nonlinear analysis	Mark P. Taylor, Lucio Sarno, Eric Girardin	<i>Empirical Economics</i>
2016	An analysis of asymmetric consumer price responses and asymmetric cost pass-through in the French coffee market	Sofia B. Villas-Boas, Céline Bonnet	<i>European Review of Agricultural Economics</i>
2008	Health information and substitution between fish: Lessons from laboratory and field experiments	Sandrine Blanchemanche, Stephan Marette, Jutta Roosen	<i>Food Policy</i>
2016	Value of parsimonious nutritional information in a framed field experiment	Jisung Jo, Jayson L. Lusk, Bernard Ruffieux, Laurent Muller	
2013	Hiring discrimination based on national origin and religious closeness: results from a field experiment in the Paris area	Guillaume Pierne	<i>IZA Journal of Labor Economics</i>
2008	On the relational motive for volunteer work	François-Charles Wolff, Lionel Prouteau	<i>Journal of Economic Psychology</i>
2005	The buyer's option in multi-unit ascending auctions: the case of wine auctions at Drouot	William Roos, Michael Visser, Philippe Février	<i>Journal of Economics and Management Strategy</i>
2009	Innovation and productivity in knowledge intensive business services	Jean-Pierre Huiban, Antonio Musolesi	<i>Journal of Productivity</i>
2013	Using money to motivate both 'saints' and 'sinners': a field experiment on motivational crowding-out	Charles Figuières, Antoine Beretti, Gilles Grolleau	<i>Kyklos</i>
2007	The effects of age and family constraints on gender hiring discrimination: a field experiment in the French financial sector	Pascale Petit	<i>Labour Economics</i>
2010	Education and early career outcomes of second-generation immigrants in France	François Poinas, Christian Belzil	
2010	Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance	Céline Bonnet, Pierre Dubois	<i>RAND Journal of Economics</i>
2012	Stability of risk preference measures: results from a field experiment on French farmers	Stéphane Couture, Arnaud Reynaud	<i>Theory and Decision</i>

2. Second test pour la France : l'identification des articles dits faux négatifs

L'identification des « faux négatifs » représente un travail considérable : il s'agit de parcourir les numéros des revues susceptibles de publier des études sur la France afin de repérer des articles d'évaluations d'impact qui ne sont pas présents dans notre base. Pour effectuer ce test dans des délais raisonnables, nous avons choisi de le restreindre aux neuf revues françaises qui sont à l'origine d'au moins deux articles d'évaluation d'impact sur la France dans notre base, et de commencer la recherche à partir de la date de versement dans

EconLit ou du début des années 2000 (étant donné que la production française est quasiment inexistante avant cette date).

L'encadré 2 ci-dessous présente la liste des revues et des numéros considérés. Au total, environ **5 300 articles** ont été testés.

Encadré 2 : Liste des revues testées

Annales d'économie et de statistique (15 articles dans la base) : 124 numéros entre 1986 et 2016, environ 900 articles examinés

Économie et prévision (4 articles dans la base) : 89 numéros entre 1996 et 2016, environ 900 articles examinés

Économie et statistique (16 articles dans la base) : 110 numéros entre 2005 et 2016, environ 500 articles examinés

Économie publique (3 articles dans la base) : 13 numéros entre 2003 et 2016, environ 100 articles examinés

Revue d'économie politique (4 articles dans la base) : 96 numéros entre 2001 et 2016, environ 800 articles examinés

Revue d'économie régionale et urbaine (3 articles dans la base) : 80 numéros entre 2001 et 2016, environ 500 articles examinés

Revue économique (14 articles dans la base) : 95 numéros entre 2001 et 2016, environ 1 000 articles examinés

Revue française d'économie (6 articles dans la base) : 74 numéros entre 1998 et 2016, environ 400 articles examinés

Travail et emploi (6 articles dans la base) : 40 numéros entre 2007 et 2016, environ 200 articles examinés

Le tableau 2 présente les articles ayant échappé à notre ciblage appartenant à la catégorie des faux négatifs. Au total, on décompte 54 faux négatifs. La raison principale réside dans le fait que les résumés des articles ne contiennent pas d'information sur la méthodologie employée, ce qui ne permet pas de les détecter ; d'autres articles disparaissent lors de nos traitements sur la base car la métadonnée « pays » n'est pas renseignée dans EconLit.

Tableau 2 : Articles « faux négatifs » de la base France identifiés sur 9 revues françaises, classés par revue de publication

Année	Titre de l'article	Auteurs	Titre du journal
2003	Comparaison de l'accès au premier emploi des apprentis et des lycéens	Liliane Bonnal, Sylvie Mendes, Catherine Sofer	<i>Annales d'économie et de statistique</i>
2010	Reconciling work and family life: the effect of the French paid parental leave	Julie Moschion	
2007	Les déterminants locaux du chômage en région parisienne	Laurent Gobillon, Harris Selod	<i>Économie et prévision</i>
2009	Apparent returns on training in French firms: impact on productivity and wages	Patrick Aubert, Bruno Crépon, Philippe Zamora	
2009	The impact of working conditions on sick leave: the case of non-standard working hours	Cédric Afssa, Pauline Givord	
2013	Perceived job security: the effects of unemployment insurance and labour court activity	Henri Fraise, Corinne Prost, Laurence Rioux	
2014	Randomised controlled trial vs. matching: the evaluation of a French job search assistance programme for young job seekers	Marianne Bléhaut, Roland Rathelot	
2014	Do job seekers benefit from more frequent interviews with their caseworkers?	Maëlle Fontaine, Thomas Le Barbançon	
2008	La norme ISO 14001 est-elle efficace ? Une étude économétrique sur l'industrie française	Nicolas Riedinger, Céline Thévenot	<i>Économie et statistique</i>
2009	Les effets des allègements de cotisations sociales sur l'emploi et les salaires : une évaluation de la réforme de 2003	Matthieu Bunel, Fabrice Gilles, Yannick L'Horty	
2011	La réforme des retraites de 1993 : l'impact de l'augmentation de la durée d'assurance	Antoine Bozio	
2012	L'apprentissage : un impact positif sur la réussite scolaire des niveaux V	Elodie Alet, Liliane Bonnal	
2014	L'impact de la participation aux pôles de compétitivité sur les PME et les ETI	Christophe Bellégo, Vincent Dortet-Bernadet	
2015	Quels effets du recrutement en contrat aidé sur la trajectoire professionnelle ? Une évaluation à partir du Panel 2008	Isabelle Benoteau	
2016	L'influence de la santé mentale déclarée sur le maintien en emploi	Thomas Barnay, Eric Defebvre	
2016	Obésité et marché du travail : les impacts de la corpulence sur l'emploi et le salaire	Elise Coudin, Arthur Soulette	
2003	Fiscalité et mariage	Sophie Buffeteau, Damien Echevin	<i>Économie publique</i>
2012	Impact d'une sur-assurance santé sur le reste à charge des patients	Aurélie Pierre, Anne Goselin	
2012	The effect of public subsidies for formal care on the care provision for disabled elderly people in France	Roméo Fontaine	

2002	La prévention du tabagisme des adolescents français	Fabrice Etilé	<i>Revue d'économie politique</i>
2004	Are R&D subsidies a substitute or a complement to privately funded R&D?	Emmanuel Duguet	
2009	La présence syndicale réduit-elle la discrimination salariale à l'encontre des femmes ? Un examen sur l'année 2002	Emmanuel Duguet, Pascale Petit	
2012	The effect of the incremental R&D tax credit on the private funding of R&D an econometric evaluation on French firm level data	Emmanuel Duguet	
2014	The effect of non-work related health events on career outcomes: an evaluation in the French labor market	Emmanuel Duguet, Christine Le Clairche	
2015	Grèves et productivité du travail : application au cas français	Jeremy Tanguy	
2007	Les rendements de la mobilité spatiale des jeunes actifs. Une analyse comparative par niveau de formation	Marie-Benoît Magrini	<i>Revue d'économie régionale et urbaine</i>
2010	Caractéristiques communales et retour à l'emploi. Une analyse empirique sur données françaises	Oana Calavrezo, Florent Sari	
2015	Home-ownership and unemployment: a French test of the Oswald hypothesis	Florent Sari	
2015	Mesure de l'impact d'une politique publique visant à favoriser l'installation et le maintien de médecins généralistes : l'exemple du soutien au développement des maisons et pôles de santé en France	Guillaume Chevillard, Julien Mousques, Véronique Lucas-Gabrielli, Yann Bourgueil, Stéphane Rlcan, Gérard Salem	
2016	L'environnement institutionnel a-t-il un impact sur les pratiques d'irrigation ?	Liliane Bonnal, Ornella Boutry	
2004	Le statut résidentiel affecte-t-il la durée de chômage ?	Carole Brunet, Jean-Yves Lesueur	<i>Revue économique</i>
2005	Évaluation des effets de l'intéressement Aubry sur l'activité des bénéficiaires des minima sociaux à la Réunion	Nadia Alibay, Nathalie Picard, Alain Trannoy	
2006	Quels effets des réorganisations sur la date de passage aux 35 heures ?	Fabrice Gilles	
2009	L'influence de l'offre de soins et du niveau des primes sur la demande d'assurance complémentaire santé en France	Renaud Legal	
2009	Les transmissions d'entreprise héréditaires sont-elles moins efficaces ?	Laurent Bach	
2009	L'effet de l'ARTT sur le chômage partiel	Oana Calavrezo, Richard Duhautois, Emmanuelle Walkowiak	
2010	La date de naissance influence-t-elle les trajectoires scolaires et professionnelles ?	Julien Grenet	
2011	La dissuasion des homicides volontaires	Nicolas Gérard Vaillant, Benoît Dervaux	
2012	Concilier vie familiale et vie professionnelle. L'effet de la préscolarisation	Julie Moschion	
2013	Le prêt à la création d'entreprises est-il efficace ?	Françoise Bastié, Sylvie Cieply	

2014	L'activité réduite des demandeurs d'emploi	Maëlle Fontaine, Julie Rochut	<i>Revue économique</i>
2015	Does gender affect scientific productivity?	Jacques Mairesse, Michele Pezzoni	
2016	Une évaluation de l'impact de l'aménagement des conditions de travail sur la reprise du travail après un cancer	Emmanuel Duguet, Christine Le Clairche	
2016	La réponse graduée de l'Hadopi a-t-elle eu des effets sur le piratage de musique et de films ?	Eric Darmon, Sylvain Dejean, Thierry Pénard	
2016	Le niveau de chômage dans le voisinage affecte-t-il l'entrée sur le marché du travail ?	Matthieu Solignac, Maxime To	
2016	Évaluation de l'effet d'une politique spatialisée d'accès au logement	Laurent Gobillon, Benjamin Vignolles	
1999	Réduction du temps de travail et emploi : quelques leçons du passage aux 39 heures de 1982	Bruno Crépon, Francis Kramarz	<i>Revue française d'économie</i>
2012	Une évaluation expérimentale d'un microprogramme social	Yannick L'Horty, Emmanuel Duguet, Pascale Petit	
2013	Évaluer un dispositif sectoriel d'aide à l'emploi : l'exemple des hôtels-café-restaurants	Mathieu Bunel	
2013	L'accompagnement personnalisé des demandeurs d'emploi	Luc Behaghel, Bruno Crépon, Marc Gurgand, Thierry Kamionka, Laurent Lequien, Roland Rathelot, Philippe Zamora	
2016	Impact des droits de succession sur le comportement d'accumulation du patrimoine	Jonathan Goupille-Lebret, Arturo Infante	
2016	Évaluer l'effet local d'une innovation : une application au marché du poisson français	Laurent Gobillon, François-Charles Wolff	
2011	Le développement de l'apprentissage dans le supérieur : une évaluation empirique à partir de l'enquête Génération 2001	Sabina Issehnane	<i>Travail et emploi</i>
2013	L'accompagnement et les trajectoires d'insertion des jeunes bénéficiaires d'un contrat d'insertion dans la vie sociale	Bernard Gomel, Sabina Issehnane, François Legendre	

3. Analyse de sensibilité des classements

Nous utilisons les tests effectués sur la France pour analyser la sensibilité de nos classements aux biais mis en évidence. Pour cela, nous opérons une distinction entre, d'une part, les revues internationales en langue anglaise et, d'autre part, les revues nationales dans une langue autre que l'anglais. S'agissant des faux négatifs, et donc de la sous-estimation de la production, il apparaît probable que ce biais affecte davantage les revues nationales que les revues internationales, dont à la fois le bon référencement dans EconLit et les standards de publication permettent un ciblage plus efficace des articles d'évaluation d'impact.

Nous construisons ensuite deux scénarios en corrigeant le nombre d'articles dans chaque pays de l'écart entre les faux négatifs et les faux positifs observés sur la base France.

A – Scénario 1 : correction restreinte à la sous-estimation du nombre d'articles publiés dans les revues nationales

Dans ce scénario, nous faisons l'hypothèse que le biais (la différence entre faux négatifs et faux positifs) porte sur les revues nationales et conduit à une sous-estimation du nombre d'articles. Pour les revues internationales, nous supposons que les erreurs se compensent et donnent un biais proche de zéro.

Notre démarche est alors de corriger l'estimation pour tous les pays dont la langue officielle n'est pas l'anglais (France, Allemagne, Pays-Bas, Suède), en appliquant à la part de leur production issue des revues nationales un coefficient qui correspond au taux de sous-estimation évalué sur la base France. Ainsi, pour la base France, dans les revues nationales, 54 « faux négatifs » et 16 « faux positifs » ont été identifiés, soit une différence de 38 articles, pour un total d'articles en langue nationale de 75. On ajoute donc ces 38 articles à la base France et on applique le même rapport aux articles en langue nationale des autres pays¹. Le tableau 3 ci-dessous présente ce calcul en volume de production, les nouveaux rangs obtenus et le coefficient multiplicateur (c'est-à-dire le taux de sous-estimation) qui s'applique finalement à la base du pays concerné :

Tableau 3 : Estimation du biais sur le volume, scénario 1

<i>Volume total de production</i>	<i>Estimation initiale</i>	<i>Rang initial</i>	<i>Nombre d'articles dans des revues en langue nationale hors anglais</i>	<i>Estimation scénario 1</i>	<i>Rang scénario 1</i>	<i>Coefficient multiplicateur</i>
États-Unis	2884	1	0	2884	1	1
Royaume-Uni	374	2	0	374	3	1
Allemagne	373	3	56	402	2	1,08
France	180	4	75	218	4	1,21
Canada	158	5	0	158	5	1
Suède	135	6	0	135	6	1
Pays-Bas	99	7	8	103	7	1,04

Ces calculs montrent un effet assez faible de la correction utilisée sur les volumes de production et sur les classements. Seul le classement relatif du Royaume-Uni et de l'Allemagne – qui étaient pratiquement à égalité dans notre estimation initiale – est modifié, mais leur production reste très proche et de l'ordre de deux fois celle de la France, qui demeure quatrième du classement. La sous-estimation du volume pour la base France est la plus importante, de 21 %, tandis qu'elle est inférieure à 10 % pour l'Allemagne et les Pays Bas.

¹ Comme le montre le tableau 3, ce calcul s'applique en fait surtout à la France et à l'Allemagne. Pour les Pays-Bas, seule une revue (*De Economist*) est concernée. Pour la Suède, aucune revue publiée en suédois n'apparaît dans notre base.

B – Scénario 2 : correction de la sous-estimation pour les revues nationales et d'une surestimation pour les revues internationales

Dans ce second scénario, nous opérons la même correction que dans l'option précédente, mais nous y ajoutons l'information obtenue sur les « faux positifs » sur la France issus des revues internationales. Nous faisons ainsi l'hypothèse que le nombre de « faux négatifs » des revues internationales est nul², et qu'il faut donc corriger pour une surestimation du nombre d'articles issus de ces revues en raison de l'existence de « faux positifs ».

Le premier test réalisé sur la France a fait apparaître 18 « faux positifs » issus de revues internationales en langue anglaise, pour un total de 105 articles. Nous soustrayons donc ces articles de la base France et appliquons le même rapport aux articles issus de revues internationales en langue anglaise dans tous les pays de notre base. Le tableau 4 ci-dessous présente ce calcul en volume de production, les nouveaux rangs obtenus et le coefficient multiplicateur (c'est-à-dire le taux de sous-estimation ou de surestimation) qui s'applique finalement à la base du pays concerné :

Tableau 4 : Estimation du biais sur le volume, scénario 2

<i>Volume total de production</i>	<i>Estimation scénario 1</i>	<i>Rang initial</i>	<i>Nombre d'articles dans des revues internationales</i>	<i>Estimation scénario 2</i>	<i>Rang scénario 2</i>	<i>Coefficient multiplicateur</i>
États-Unis	2884	1	2884	2394	1	0,83
Royaume-Uni	374	3	374	310	3	0,83
Allemagne	402	2	317	348	2	0,93
France	218	4	105	200	4	1,11
Canada	158	5	158	131	5	0,83
Suède	135	6	135	112	6	0,83
Pays-Bas	103	7	91	88	7	0,88

Les résultats de ce deuxième calcul sont sensiblement les mêmes que ceux du scénario 1 en termes de classements, quoique l'application d'un coefficient de surestimation aux revues en langue anglaise amène à réduire le volume de production pour tous les pays excepté la France. Le Royaume-Uni et l'Allemagne voient leur classement relatif s'inverser un peu plus nettement. La France reste placée en quatrième position. Elle fait l'objet d'une sous-estimation de 11 %, tandis que les autres pays font l'objet d'une surestimation comprise entre 7 % et 17 %.

² Idéalement, pour calculer ce scénario, il faudrait pouvoir estimer également le taux de faux négatifs à appliquer aux revues internationales, à partir d'un échantillonnage de revues similaire à celui effectué sur les revues françaises. Pour des raisons de faisabilité, nous n'avons pas réalisé une telle estimation.

C – Synthèse

Pour conclure ces tests de sensibilité, nous reportons dans le tableau 5 ci-dessous les résultats obtenus lorsqu'on applique les taux de sous-estimation ou de surestimation issus des deux scénarios précédents à la production par million d'habitants des pays de notre échantillon.

Tableau 5 : Résultats pour la production rapportée à la population

<i>Production par million d'habitants</i>	Estimation initiale	Rang initial	Estimation scénario 1	Rang scénario 1	Estimation scénario 2	Rang scénario 2
Suède	13,91	1	13,91	1	11,55	1
États-Unis	9,03	2	9,03	2	7,49	2
Pays-Bas	5,87	3	6,11	3	5,19	3
Royaume-Uni	5,81	4	5,81	4	4,82	4
Allemagne	4,63	5	4,98	5	4,32	5
Canada	4,44	6	4,44	6	3,69	6
France	2,72	7	3,30	7	3,02	7

Il en ressort que les résultats en termes de volume de production rapporté à la population ne sont pas affectés par les biais estimés. La Suède reste le pays le plus productif, suivie d'assez loin par les États-Unis et les Pays-Bas. Le Royaume-Uni et l'Allemagne ont des productions relativement comparables. La France est placée loin derrière les autres pays, y compris le Canada.

RETROUVEZ
LES DERNIÈRES ACTUALITÉS
DE FRANCE STRATÉGIE SUR :



www.strategie.gouv.fr



[francestrategie](https://www.facebook.com/francestrategie)



[@Strategie_Gouv](https://twitter.com/Strategie_Gouv)



[france-strategie](https://www.linkedin.com/company/france-strategie)



[@francestrategie_](https://www.instagram.com/francestrategie_)



FRANCE STRATÉGIE

France Stratégie est un organisme d'études et de prospective, d'évaluation des politiques publiques et de propositions placé auprès du Premier ministre. Lieu de débat et de concertation, France Stratégie s'attache à dialoguer avec les partenaires sociaux et la société civile pour enrichir ses analyses et affiner ses propositions. Elle donne à ses travaux une perspective européenne et internationale et prend en compte leur dimension territoriale.