UNIVERSITÉ D'AIX-MARSEILLE

FACULTÉ DES SCIENCES ECONOMIQUES ET DE GESTION

Ecole Doctorale de Sciences Economiques et de Gestion d'Aix-Marseille n° 372

Année 2012 Numéro attribué par la bibliothèque

| | | | | | | | | | | | |

Thèse pour le Doctorat ès Sciences Economiques

Présentée et soutenue publiquement par

Clément BOSQUET

le 3 octobre 2012

COMMERCE INTERNATIONAL ET ÉCONOMIE DE LA SCIENCE : DISTANCES, AGGLOMÉRATION, EFFETS DE PAIRS ET DISCRIMINATION

JURY

RAPPORTEURS

M. Gilles DURANTON	Professeur à l'Université de Toronto
M. Thierry MAYER	Professeur à Sciences-Po

EXAMINATEURS

M. Yann BRAMOULLÉ	Professeur à l'Université de Laval
M. Pierre-Philippe COMBES	Directeur de Recherche CNRS (directeur), GREQAM
M. Juan José DOLADO	Professeur à l'Université Carlos III de Madrid
Mme. Cecilia GARCIA-PEÑALOSA	Directeur de Recherche CNRS au GREQAM
M. Thierry MAGNAC	Professeur à l'Université de Toulouse 1

AIX-MARSEILLE UNIVERSITY

Faculté des Sciences Economiques et de Gestion

Ecole Doctorale de Sciences Economiques et de Gestion d'Aix-Marseille n° 372

Year 2012 Numéro attribué par la bibliothèque

| | | | | | | | | | | |

Thesis to obtain the title of Doctor in Economics

Presented and publicly defended by

Clément BOSQUET

October, 3rd, 2012

International trade and economics of science : distances, agglomeration, peer effects and discrimination

Jury

Referees

Mr. Gilles DURANTON	Professor, University of Toronto
Mr. Thierry MAYER	Professor, Sciences-Po

Examiners

Mr. Yann BRAMOULLÉ	Professor, Université Laval
Mr. Pierre-Philippe COMBES	CNRS Research Professor (advisor), AMSE
Mr. Juan José DOLADO	Professor, Universidad Carlos III de Madrid
Mrs. Cecilia GARCIA-PEÑALOSA	CNRS Research Professor, AMSE
Mr. Thierry MAGNAC	Professor, University of Toulouse 1

Remerciements et pensées

Au crépuscule d'une telle aventure, il est d'usage de commencer par remercier son directeur de thèse... Mais si je me joins à cette tradition aujourd'hui, ce n'est absolument pas par obligation mais parce que Pierre-Philippe a su orienter ces quatre années vers un épanouissement professionnel total grâce à sa disponibilité, son exigence, sa transparence et son investissement. Le tout, toujours dans le bonne humeur, il a créé une ambiance de travail des plus agréables et des plus motivantes. Si je ne garde que de bons souvenirs de mes années marseillaises et si je suis enfin sûr d'avoir trouvé ma voie, c'est grâce à toi, chef.

Comme deux époux doivent bénir leur entremetteur, je remercie également avec grand plaisir Thierry Mayer de m'avoir guidé vers Pierre-Philippe au moment du choix d'un directeur de thèse. Merci aussi d'avoir encouragé mes premières recherches lorsque j'étais en Master et d'avoir accepté d'être rapporteur de cette thèse.

J'en profite pour remercier chaleureusement mes deux rapporteurs, Gilles Duranton et Thierry Mayer donc, ainsi que Thierry Magnac pour leurs précieux conseils délivrés (à l'écrit ou à l'oral) lors de la pré-soutenance. Ils ont largement contribué à l'amélioration de ce manuscrit et des articles qui en découleront.

Un grand merci également à Cecilia Garcia-Peñalosa pour avoir accepté de faire partie de mon jury mais surtout pour ses conseils diffusés avant et pendant notre passionnante collaboration. Enfin, merci à Juan José Dolado et à Yann Bramoullé de venir de si loin (sic !) pour également participer au jury de cette thèse.

Parmi ceux qui ne seront pas à ma soutenance, je tiens à remercier particulièrement Hervé Boulhol, mon coauteur depuis cinq ans déjà (courage, on va les terminer ces papiers !)

pour m'avoir également donné goût à la recherche pendant mon Master. Je pense également à cet instant à Michel Fouquin, mon premier coauteur quand j'étais stagiaire au CEPII. Merci aussi à Laurent Gobillon pour ses idées précieuses et ses apports économétriques non négligeables.

Pour leur accueil chaleureux, leurs conseils avisés et les excellentes conditions de travail, je tiens à remercier Bruno Decreuse, Tanguy van Ypersele, Russell Davidson, Alain Trannoy, Sebastian Bervoets, Frederic Deroian, Jean-Benoît Zimmerman et Nicolas Gravel du GREQAM ainsi que Thierry Mayer, Jean-Marc Robin, Philippe Martin, Etienne Wasmer, Guy Laroque, Yann Algan, Koen Jochmans, Ruben Durante, Alexandre Biotteau et Guillaume Sarrat de Tramezaigues de Sciences Po.

Merci également à tout ceux qui, accompagnés de leur gentillesse, nous facilitent le travail au quotidien, je pense en particulier à Bernadette, Corinne, Isabelle, Diane et Yves au GREQAM et à Cécile, Pilar et Cathy à Sciences Po.

A la fin de cette aventure, j'ai une pensée pour mes amis économistes qui l'ont accompagnée. D'abord à Marseille, comment ne pas commencer par Benoit, ami des plus hautes cimes et camarade en (presque) toutes occasions, quel plaisir de t'avoir aussi souvent vu à la maison et ailleurs. Embrasse Cécile et Lucien pour moi, amusez vous bien à Washington. Après un tel éloge, je ne terminerais pas ma soutenance vivant si je ne citais pas immédiatement Momo, capable de m'envoyer ses amis de la pègre si elle est vexée, et qui conserve toute mon affection malgré son sale caractère ! Un grand merci également à Kalila pour son formidable accueil et l'amitié qui en a suivi. Mes autres co-bureaux, Maame, Shamaila, Natalyia, Chen et Vincent et certains amis sociologues ne me démentiront pas si je dis qu'elle est une pièce maitresse de la vie à la Vieille Charité. Toujours ravi de profiter des

fantaisies et de la bonne humeur de Sarra, j'ai également une pensée pour les plus vieux qui m'ont baptisé (...) quand je suis arrivé au GREQAM : Ophélie, Paul, Renaud et Mathieu. Zak, camarade de footing et de bien d'autres choses, Luis, Rémy, Maty et Aziz ne sont également pas à oublier.

Merci également à Marion même si elle joue foireusement à la Koinche, et à mes amis sociologues marseillais, Tanguy et Sophia, Julie, JB et Jean-Christophe, Justine, Chiara, tous pour tant de choses beaucoup trop longues (et personnelles !) à expliquer ici. Ma soutenance qui aura lieu à Marseille sera, je l'espère, l'occasion de se remémorer tous ces excellents souvenirs. Rencontrés ou retrouvés au hasard de la vie marseillaise, je pense également à Marie, Claire, Léa, Sarah et Yann.

Un immense merci également à Edith, Guillaume, Clarisse, Lucas et Léonard pour leur fantastique accueil et grâce à qui j'ai été enchanté de vivre dans le quartier Notre-Dame-du-Mont à Marseille.

Puis l'aventure s'est poursuivie à Sciences Po, une pensée particulière alors pour Camille et ses soirées épiques, Dilan, G.O. Marion, Nina, Lisa, François et ses chemises, Jamil, Jules, Camilo et Marc. En revenant sur Paris, j'ai pu également y voir plus souvent des amis économistes de longues dates, Clémence, Charlotte, Hélène, Alexandre, Clément, Mathieu, Marc (again !) et de mois longue date, Sophie, Fabienne, Mathieu...

Pendant tout ce temps, je n'ai pas perdu contact avec mes camarades les plus fidèles qui vont me manquer dans les années à venir, Clémclem, Bourboussette et Djidji (trop à dire sans s'éterniser), Beber - caviste adoré - et MiniBlonde, Ilinca et Adrien (parle moins vite !), Q et Fleu, Raoul et Flo, Toto et Virg', Julie et Mathias (félicitations !), Fanny-Gégé, Marie-Francis, Etienne Hey Beurkel !, Aurore Hey Sccchhhiltte !, Hélène, Ysa, Nabil, Léo, Mag et Colinou, Amélie et Filou, Marc et Aurore et Alex et Guysèle chez qui je ne manquerai pas d'aller boire le thé pendant les week-ends des années à venir. Une pensée des plus tendres pour mes amis d'enfance, Cloé, Camille, Marie, Agnès, Eloi, Thomas et Olivier -futur voisin !-, ceux du lycée, Adrien et Cyrielle, Valérie et Guillaume ainsi que Marie et Matthieu à qui je souhaite le meilleur pour les années à venir, et à ceux que la vie rend difficile à mettre dans des cases, ma colocataire Maureen, Dave (et crac !) et Claire.

Enfin, j'ai une pensée particulière pour ma famille proche, mes parents qui ont su me soutenir en toutes circonstances et surtout me pousser quand c'était nécessaire. Sans vous je n'en serais certainement pas là aujourd'hui. Olivia, ma frangine adorée sait déjà tout le bien que je pense d'elle, de Thomas et Capucine, mais rares sont les occasions d'écrire à nos proches à quel point on les aime. Et Rom, comme tous les frères on s'est battu quand on était gamins, mais quelle joie d'avoir aujourd'hui cette relation épanouie avec toi ! Commence à préparer ta soutenance, c'est toi le prochain ! Je pense aussi à ma famille au sens large, à mes grands parents. Il serait trop long de citer tous les oncles et tantes, cousins et cousines. Mention spéciale donc à ceux de ma génération, Greg, Marie, Peggy et Alice, aux frangines Lemaires ainsi qu'à Marie-Françoise et Jean-Michel pour conserver ma cave à température optimale pendant mes années londoniennes !

Comme toujours en de pareilles circonstances, j'espère n'avoir oublié personne. Si c'était le cas (et même si ça ne l'est pas), je m'en excuse profondément et vous invite à me rendre visite à Londres pour me faire pardonner.

Résumé

Cette thèse rassemble principalement des contributions en économie de la science à laquelle les deux premières parties sont consacrées. La première teste l'importance des choix méthodologiques dans la mesure de la production scientifique et étudie les canaux de diffusion de la connaissance. La deuxième s'intéresse aux déterminants individuels et locaux de la productivité des chercheurs et au différentiel de promotion entre hommes et femmes sur le marché du travail académique.

Sont établis les résultats suivants : les choix méthodologiques dans la mesure de la production scientifique n'affectent que très peu les classements des institutions de recherche. Les citations et les poids associés à la qualité des journaux mesurent globalement la même productivité de la recherche. La localisation des chercheurs a un impact sur leur productivité dans la mesure où certaines universités génèrent davantage d'externalités que d'autres. Ces externalités sont plus importantes là où les chercheurs sont homogènes en terme de performances, où la diversité thématique est grande, et dans une moindre mesure dans les grands centres de recherche, lorsqu'il y a plus de femmes, de chercheurs âgés, de stars et là où les chercheurs sont connectés à des co-auteurs à l'étranger. Si les femmes sont moins souvent Professeur des Universités (par opposition à Maître de Conférences) que les hommes, ce n'est ni parce qu'elles sont discriminées dans le processus de promotion, ni que le coût de promotion (mobilité) est plus important pour elles, ni qu'elles ont des préférences différentes concernant le salaire et le prestige des institutions dans lesquelles elles travaillent. Une explication possible, mais non testée, serait donc que les femmes s'autocensurent ou exercent moins d'efforts que les hommes pendant les processus de promotion.

La troisième partie de cette thèse contribue à la littérature sur le choix des estimateurs pour les équations de gravité. J'y discute d'abord une hypothèse forte des articles remettant en cause les estimations par moindres carrés ordinaires du modèle log-linéarisé. Puis l'utilisation des nouveaux estimateurs est testée dans l'analyse du "paradoxe de la distance". Enfin, il est montré que les estimateurs de pseudo-maximum de vraisemblance à partir d'une loi binomiale négative utilisés dans la litérature dépendent de l'unité de mesure de la variable dépendante. Un nouvel estimateur surmontant cette limite est présenté.

Mots clés : économie de la science, production scientifique, citations, publications, déterminants de la productivité, effets de pairs, discrimination, équations de gravité, méthodes de pseudo-maximum de vraisemblance

Codes JEL : I23, R12, J24, O31, J45, C13

Abstract

The core of this thesis lies in the field of economics of science to which the two first parts are devoted. The first part questions the impact of methodological choices in the measurement of research productivity and studies the channels of knowledge diffusion. The second part studies the impact on individual publication records of both individual and departments' characteristics and analyse the gender gap in occupations on the academic labour market.

The main results are the following: methodological choices in the measurement of research productivity do not impact the estimated hierarchy of research institutions. Citations and journal quality weights measure the same dimension of publication productivity. Location matters in the academic research activity: some departments generate more externalities than others. Externalities are higher where academics are homogeneous in terms of publication performance and have diverse research fields, and, to a lower extent, if the department is large, with more women, older academics, stars and co-authors connection to foreign departments. If women are less likely to be full Professor (with respect to Assistant Professor) than men, this is neither because they are discriminated against in the promotion process, neither because the promotion cost (mobility) is higher for them, nor because they have different preferences for salaries versus department prestige. A possible, but not tested, explanation is that women self-select themselves by participating less in or exerting lower effort during the promotion process.

The third part of this thesis deals with a completely different literature: the choice of estimators for international trade gravity equations. First, I discuss an important assumption of articles contesting the use of ordinary least squares of log-linearised models. Then, the "distance puzzle" is re-analysed by testing the use of the proposed non-linear estimators in levels. Last, Negative Binomial Pseudo-Maximum Likelihood estimators used in the literature are scale-dependent. A new version of this estimator overcoming that limitation is introduced.

Key words: economics of science, research productivity, citations, publications, productivity determinants, peer effects, discrimination, gravity equations, Pseudo-Maximum Likelihood methods

JEL Classification: I23, R12, J24, O31, J45, C13

Avertissement

Mis à part l'introduction, le premier chapitre et la conclusion de cette thèse, les chapitres 2 à 7 sont issus d'articles de recherche rédigés en anglais et dont la structure est autonome. Ceci y explique la présence des termes “paper” ou “article” ainsi que l'éventuelle répétition de certaines informations. Les chapitres 1 à 4, 6 et 7 sont issus de collaborations avec mes coauteurs, ce qui y justifie lŠutilisation des pronoms “we” ou “nous”.

Notice

Except the general introduction, the first chapter and the conclusion, remaining chapters (2 to 7) of this thesis are self-containing research articles. Consequently, terms “paper” or “article” are frequently used. This also explain that some information are given in multiple places of the thesis. Chapters 1 to 4, 6 and 7 are written with co-authors, what explain the use of the “we” or “nous” (in french) pronouns.

Table des matières

Liste des tableaux

Table des figures

Introduction générale

Cette thèse rassemble majoritairement des contributions en économie de la science. L'introduction est rédigée principalement dans cette direction.

Contexte

Quelle que soit l'opinion qu'on peut porter sur les réformes de l'enseignement supérieur et de la recherche engagées par différents gouvernements quasi-successifs au milieu des années 2000, il est indéniable qu'elles ont eu des répercussions importantes sur le monde universitaire en France. En 2005, le gouvernement Jean-Pierre Raffarin crée l'Agence Nationale de la Recherche qui est une agence de moyens finançant directement les équipes de recherche, par projets concurrents, sous forme de contrats de courtes durées. En 2006, la loi de programme pour la recherche, promulguée par le gouvernement Villepin, crée l'AERES (Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur) chargée d'évaluer à la fois les établissements publics d'enseignement supérieur et de recherche et les formations et diplomes associés qu'ils délivrent, mais aussi de définir les conditions de l'évaluation des chercheurs.[1] En 2007, Valérie Pécresse alors ministre de l'enseignement supérieur et de la

1. Loi n° 2006-450 du 18 avril 2006 de programme pour la recherche.

recherche fait promulguer par le gouvernement Fillon la loi relative aux libertés et responsabilités des universités (dite loi LRU) dont les décrets d'application consacrent l'autonomie des universités dans les domaines budgétaires, de gestion des ressources humaines et de propriété de leurs biens immobiliers et réforment certaines procédures principalement en matière de gouvernance et de recrutement.[2] Finalement, le grand emprunt de 2010, baptisé "investissements d'avenir" finance notamment le Plan Campus, ayant pour objectif de faire émerger une douzaine de pôles universitaires d'excellence au niveau international, et la création de LabEx (Laboratoires d'Excellence), d'EquipEx (Equipements d'Excellence) et d'IDEx (Initiatives d'Excellence), grâce à des dotations exceptionnelles. Mis en place en 2011-2012, les LabEx ont ainsi vocation à faire jeu égal avec leurs homologues internationaux en ayant une visibilité internationale construite grâce à une politique de recherche et de formation de haut niveau, attirant des (enseignants-)chercheurs de renommée mondiale à l'aide de ces nouveaux moyens.

Ainsi, le milieu universitaire français, qui fait face à une concurrence internationale de plus en plus grande, notamment au niveau de la recherche, a connu des évolutions institutionnelles importantes ces dernières années. Cette thèse n'a pas pour objet d'essayer d'évaluer ces réformes, ce qui nécessiterait, d'une part, d'en connaître les objectifs précis en termes de performances économiques et, d'autre part, davantage de recul temporel afin de bénéficier de données statistiques suffisantes. Cependant, ce contexte politique est propice à vouloir proposer des avancées significatives sur quelques points peu étudiés de la litérature : comment évaluer les institutions de recherche ? Comment organiser l'activité de recherche

2. Loi n° 2007-1199 du 10 août 2007 relative aux libertés et responsabilités des universités (dite loi LRU ou loi Pécresse), initialement intitulée loi portant organisation de la nouvelle université et communément appelée loi d'autonomie des universités.

universitaire ? Quel design institutionnel est par exemple souhaitable pour les laboratoires publiques ?

L'économie de la science et de l'innovation

Ces vastes interrogations nous ont conduit vers un questionnement scientifique naturellement plus précis : quelle est l'importance des choix méthodologiques sur les évaluations et la hierarchie estimée des institutions scientifiques ? Comment est diffusée la connaissance ? Quelle est l'organisation optimale de l'activité de recherche universitaire ? Et, dans une autre dimension, que peut-on dire des écarts de promotion entre hommes et femmes sur ce marché du travail particulier qu'est le marché du travail académique ?

La majorité de cette thèse rassemble donc des contributions en économie de la science et de l'innovation. Dans le New Palgrave Dictionary of Economics, Arthur M. Diamond, Jr. (2007) définit trois objectifs au champ de l'économie de la science : comprendre l'impact de la science sur les avancées technologiques, expliquer les comportements des scientifiques et analyser l'efficacité ou l'inefficacité des institutions scientifiques. En outre, ce champ est considéré comme important à travers les impacts successifs de la science sur la technologie et de la technologie sur la productivité et la croissance. L'influence de ce champ est également grandissante dans un contexte où les débats sur les limites finies de la planète font des gains de productivité l'argument essentiel de la croissance de demain.

Cette thèse emprunte également des méthodes et des questionnements à l'économie géographique, discipline en plein essor depuis le prix Nobel de Paul Krugman en 2008 et dont les objets principaux sont l'évaluation des économies d'agglomération et des politiques régionales dont les enjeux peuvent être importants (World Development Report, 2009). Les progrès faits par les instituts statistiques de beaucoup de pays permettent d'accéder à des données dont la qualité s'améliore de plus en plus, ce qui élargie les possibilités d'identifications et clarifie les directions de causalité. Les premières parties de cette thèse incluent aussi des éléments issus de l'économie du travail dans la mesure où elles touchent aux thèmes de la productivité du travail, des effets de pairs et de réseau, et de la discrimination.

Dans le contexte politique décrit ci-dessus, cette thèse s'intéresse à la recherche publique, c'est-à-dire universitaire et assimilée (Cnrs, Inra). [3] Néanmoins, il est possible de considérer que les conclusions de cette thèse ont une portée plus générale dans la mesure où la recherche académique est souvent considérée comme un élément important de l'innovation dans les activités industrielles (Jaffe, 1989 ; Acs, Audretsch et Feldman, 1991 ; Mansfield, 1995 ; Mansfield, 1998). Plus précisément, les travaux décrits ici utilisent des données sur les économistes universitaires exerçants en France.

Pourquoi l'économie ?

S'intéresser aux économistes pourrait apparaître dans un premier temps comme n'étant pas le choix le plus naturel dans la mesure où certaines disciplines des sciences dures (chimie ou physique par exemple) sont probablement plus proches des secteurs industriels dans lesquels l'innovation est la plus grande. Ce choix a été fait pour plusieurs raisons. Premièrement, nous avons pu bénéficier d'excellentes données : la Direction Générale de la Recherche et de l'Innovation du Ministère de l'Enseignement Supérieur et de la Recherche nous a fourni la liste des enseignants chercheurs en économie travaillant en France pour chaque année entre 1990 et 2008. Cette liste aurait sans doute pu être également obtenue

3. Centre National de la Recherche Scientifique et Institut National de la Recherche Agronomique.

pour une autre discipline mais celle que nous avons obtenu a pu être fusionnée avec la base de données Econlit qui recense plus de 1200 journaux scientifiques en économie, ce qui est considérable. Deuxièmement, il existe dans la discipline une littérature assez avancée et de plus en plus acceptée sur les questions de mesure de la production scientifique et notamment de la qualité des journaux/revues. Troisièmement, les conventions de signature des articles par leurs auteurs est beaucoup plus facilement interprétable en économie que dans la plupart des sciences dures où les assistants de recherche, directeurs de thèse et de laboratoire co-signent toutes les publications. Il y est donc beaucoup plus difficile de savoir qui a contribué à tel ou tel projet de recherche et nous aurions dû alors faire des hypothèses beaucoup plus drastiques en ne considérant que les premiers auteurs par exemple. Quatrièmement, il existe en économie une classification des domaines par codes JEL (Journal of Economic Literature) qui permettent d'affiner l'analyse et d'étudier les impacts de la diversité scientifique et de la spécialisation. Cinquièmement, nous avons, en tant qu'économistes, un socle de connaissances de l'organisation du travail universitaire en économie supérieure à celles que nous aurions en étudiant une autre discipline, ce qui nous permet sans doute d'éviter certaines erreurs. Nous avons parfaitement conscience qu'aucun de ces arguments n'est décisif en lui-même ; c'est leur accumulation qui nous a conduit dans cette direction.

Cette thèse rassemble donc des contributions en économie de la science qui cherchent à repousser les frontières de la connaissance économique sur les questions de mesure de la productivité des chercheurs, des canaux de la diffusion des connaissances, sur les déterminants individuels et locaux de la productivité des chercheurs ou encore sur les différentiels de poste atteint/écarts de promotions (de niveau d'occupation) entre hommes et femmes.

En effet, travailler avec ce type de données de publications permet de construire des mesures de productivité individuelle, ce qui n'est en général pas possible sur les marchés du travail traditionnels.

Les deux premières parties de cette thèse sont donc consacrées à des questionnements en économie de la science. Plus particulièrement, le troisième chapitre étudie les déterminants locaux de la productivité des chercheurs, question qui est liée à l'organisation spatiale des activités économiques en général et de la recherche en particulier. Depuis longtemps, des recherches en économie géographique travaillent sur les liens qui existent entre l'organisation spatiale des activités et les coûts aux échanges. La dernière partie de cette thèse consiste en des développements économétriques sur ce thème, mais cette fois-ci avec des applications plus classiques sur des secteurs d'activité marchands. Elle s'intéresse notamment à l'impact de la distance géographique, comme proxy des coûts de commerce, sur le commerce international. L'accent est néanmoins mis sur les interrogations méthodologiques qui font débat actuellement pour estimer des équations de gravité.

Résumés

Les deux premières parties de cette thèse contiennent donc des essais en économie de la science alors que la troisième partie s'interroge sur le choix des estimateurs les mieux adaptés pour évaluer les paramètres d'intérêt des équations de gravité appliquées au commerce international.

Première partie

Cette partie est consacrée aux questions de la mesure de la production scientifique (chapitre 1) et aux déterminants individuels des stocks de publications et de citations (chapitre 2). Le premier chapitre n'a pas pour objet de revenir sur les bonnes ou mauvaises raisons de l'existence d'évaluations bibliométriques de la production scientifique mais de débattre de l'impact de certains choix méthodologiques relatifs aux critères d'évaluation des publications académiques en économie.

Chapitre 1

Les classements des centres de recherche et universités en économie passent la plupart du temps par une inférence indirecte de la qualité des publications via la qualité du journal dans lequel elles sont publiées (et leur longueur en nombre de pages). Le principal reproche fait à ce type d'approche est la relativement grande variabilité de qualité qu'il peut rester entre les différents articles d'un même journal, même à longueur donnée. Il est alors souvent proposé d'utiliser directement le nombre de citations reçues par chaque publication. Cela suppose tout d'abord que ce nombre de citations est un bon indicateur de la qualité de la publication, ce qui est généralement admis. Il ne faut pas oublier que cette hypothèse n'est pas complètement triviale, la littérature soulignant par exemple les fortes différences de pratique de citations entre domaines, la nécessaire pondération des citations reçues par la qualité de la publication citant, ce qui est difficile à implémenter, et le fait que le nombre de citations d'une publication est très affecté par son cycle de vie, ce qui rend crucial le choix de la fenêtre temporelle pendant laquelle les citations sont recueillies, en fonction de l'âge de la publication. Malgré ces limites, utiliser les citations de chaque publication est souvent considéré comme souhaitable mais rarement effectué du fait du nombre très limité de bases de données le permettant. La principale source, le Journal of Citation Reports (JCR) de Thomson-Reuters fournit les citations reçues par les articles publiés dans 304 journaux, alors qu'on en recense plus de 1200 dans la base de données Econlit par exemple (qui regroupe la plupart des revues de recherche en économie). Il est important de noter que la limitation est double : d'une part, une publication n'est citée que si elle appartient à un de ces 304 journaux, d'autre part, les citations ne sont recueillies que dans ces mêmes 304 journaux.[4] L'objet de ce chapitre est de déterminer si l'utilisation de Google Scholar (GS) pourrait constituer une alternative intéressante à ces deux stratégies, inférence indirecte de la qualité ou citations du JCR.

GS est un outil qui permet de considérablement dépasser les deux limites que présentent le JCR. GS recense les publications, de tout type, présentes sur Internet et calcule les citations qu'elles reçoivent sur des supports eux-mêmes présents sur Internet, de quelque nature qu'ils soient. Ainsi, est élargi à la fois le type de publications susceptibles de recevoir des citations, ce qui prend notamment en compte les supports potentiellement importants pour les économistes que constituent les ouvrages ou les documents de travail, mais élargit considérablement le nombre de supports citant. De plus, le domaine d'étude est a priori moins restreint que lorsqu'on le définit ex-ante en le réduisant aux journaux référencés par Econlit ou encore plus par le JCR. Si un économiste a une publication importante dans un journal en mathématiques non référencé par Econlit ou le JCR, celle-ci peut être considérée via GS.

4. Combes et Linnemer (2003a) comparent pour la France différentes approches qui utilisent ces citations afin de hiérarchiser l'impact des publications en économie des centres de recherche et universités françaises en 1998.

Nous proposons ici une étude prospective ayant pour objet d'étudier les propriétés d'un certain nombre d'indicateurs d'impact des centres de recherche et universités françaises en 2008 fondés sur les citations GS reçues par leurs membres en économie dans les sujets "Business, Administration, Finance, and Economics" et "Social Sciences, Arts, and Humanities". Nous calculons, sur cinq périodes de temps, cinq indicateurs de citation GS : le nombre total de citations, le nombre de citations par entrée GS, le nombre de citations par entrée GS ayant reçu au moins une citation, l'indice H, et l'indice G, tous ces indices étant dupliqués selon que l'on prend en compte ou pas le nombre de co-auteurs. Les résultats obtenus ne sont pas fondamentalement différents entre certaines de ces variantes. Ainsi, avons-nous choisi de nous concentrer dans les sections de classements, sur le nombre de citations totales et l'indice G, en volume et par chercheur, et pour une seule période donnant plus de poids aux publications récentes. Les résultats obtenus sont systématiquement comparés à ceux utilisant une approche plus traditionnelle basés sur les publications Econlit avec pondération modérée et forte de la qualité moyenne des journaux.

Les grandes tendances qui se dégagent de cette étude sont les suivantes :

i. Pour le même ensemble de chercheurs français, alors qu'Econlit recense 10154 articles (équivalent écrit seul), les sujets "Business, Administration, Finance, and Economics" et "Social Sciences, Arts, and Humanities" de GS recensent 42448 entrées dont 20934 ont reçu au moins une citation, soit un support entre deux et quatre fois plus large.

ii. Cette production accrue des français n'est pas simplement due au fait que les chercheurs publiant selon Econlit ont plus d'entrées GS que de publications Econlit mais aussi au fait que la part des non publiant est plus faible selon GS. Seuls 6.2% des chercheurs n'ont pas d'entrée GS et 15% n'en ont pas ayant au moins une citation, alors que 27.1% n'ont pas de publication Econlit. Sur les cinq dernières années, alors que 45.3% n'ont pas de publication Econlit, seulement 22.4% n'ont pas d'entrée GS et 34.3% n'en ont pas avec au moins une citation.

iii. L'ensemble des chercheurs localisés en France ont ensemble quasiment 270000 citations GS en janvier 2010, 47000 pour leurs entrées de la période 2004-2008. Chaque entrée a en moyenne presque 7 citations, le double pour celles en ayant au moins une. L'ensemble des entrées GS de la France entraîne un indice H de 180, soit 180 entrées ayant au moins 180 citations et environ 500 entrées ayant en moyenne 500 citations (indice G). Sur 5 ans, ces deux indices sont divisés par un peu plus que deux. La France obtient environ 34000 citations GS par année d'existence de chaque entrée, et une nouvelle entrée obtient en moyenne plus d'une citation par année, 2 une fois qu'elle en a obtenue au moins une.

iv. Prendre ou pas le nombre de co-auteurs ou ramener les citations par entrée ayant ou pas au moins une citation n'a qu'une influence très marginale sur les classements.

v. La corrélation élevée entre classements en nombre de citations totales et nombre de citations par entrée montre que les institutions recevant beaucoup de citations sont celles où le nombre de citations reçues par chacune des publications est elle même élevée. Recevoir beaucoup de citations ne consiste pas seulement à effectuer beaucoup de travaux mais à également avoir une qualité élevée pour chacun d'entre eux. Les indices H et G sont également très corrélés aux indices de citations totales, légèrement moins avec les indices de citations par entrée.

vi. La hiérarchie des institutions évaluée selon les approches GS, quelles qu'elles soient, est relativement proche de la hiérarchie obtenue via Econlit, avec des corrélations de rang supérieures à 0.8 que ce soit pour les citations totales reçues ou les indices H ou G. Des corrélations légèrement plus faibles sont observées avec les scores par entrées, de façon relativement naturelle puisque les scores Econlit correspondent à des volumes. Ainsi, utiliser la qualité des journaux comme prédicteur de la qualité des publications, en tous les cas du nombre de citations qu'elles reçoivent, semble constituer une stratégie raisonnable.

vii. Ces corrélations sont légèrement moins élevées en ce qui concerne la hiérarchie des institutions selon leurs scores par chercheur, ce qui pourrait s'expliquer par la plus grande variabilité globale de ceux-ci.

viii. La productivité des chercheurs pour les indices GS, quels qu'ils soient, suit une courbe en cloche au cours de leur cycle de vie, comme on l'obtient avec les scores Econlit, les chercheurs ayant un pic de productivité généralement observé entre 40 et 50 ans.

ix. Les quatre ou cinq statuts les plus productifs selon les indices GS restent identiques à ceux identifiés au moyen des scores Econlit (Ingénieurs Ponts et Chaussées, Directeur d'Etudes Ehess, Administrateurs Insee et Directeurs de recherche Cnrs). En revanche, progressent selon les mesures GS les statuts qui ne sont pas au cœur de notre analyse (hors section 5 Cnu ou section 37 Cnrs, notamment les "assimilés" professeurs ou chargés de recherche, en poste dans des grandes écoles, parfois de commerce, ou dans des administrations ou les chercheurs non section 37).[5]

x. D'une part, pour une majorité de centres ou universités, la variation de classements par rapport à Econlit est faible. D'autre part, pour quelques autres, elles est relative-

5. CNU : Conseil National des Universités, section 05 : Sciences économiques. Section 37 Cnrs : économie et gestion.

ment forte, et il s'avère que ces unités semblent être celles dont le cœur de l'activité n'est pas l'économie, ou, ce qui est lié, dont les membres ne sont pas au cœur de notre champ, à savoir les sections 5 du Cnu et 37 du Cnrs. C'est cette dernière propriété qui fait de GS un instrument complémentaire de l'instrument Econlit.

Chapitre 2

Une fois ces analyses méthodologiques réalisées, nous nous sommes intéressés aux déterminants individuels des stocks de publications et de citations des chercheurs. Nous nous sommes demandé si les déterminants standards de la productivité en économie du travail sont aussi déterminants de la productivité de la recherche, si la spécialisation disciplinaire a un impact sur la productivité des chercheurs, si les indices de citation Google Scholar et les scores de publication Econlit tenant compte de la qualité des journaux mesuraient la même productivité et comment était diffusée la connaissance. Cette dernière question, au cœur du champ de l'économie de la science, a été traitée ici en étudiant la relation économétrique entre citations et publications et en se concentrant sur les déviations qui permettaient à certains chercheurs d'être plus cités que d'autres *ceteris paribus*, c'est-à-dire avec des stocks de publications équivalents (en quantité et en qualité). Nous avons utilisé la même base de données que lors du chapitre précédent pour établir les résultats suivants :

i. La taille des équipes de collaboration, mesurée par le nombre moyen d'auteurs par article, première variable spécifique à l'activité de recherche introduite en sus des caractéristiques démographiques (âge, âge au carré et sexe) a un impact plus robuste que ces caractéristiques démographiques individuelles dans le sens où elle est un argument toujours positif et significatif des indicateurs de performances scientifiques prenant en compte la qualité d'une manière ou d'une autre (citations ou qualité des journaux).

ii. Il existe des rendements croissants de la quantité et de la taille des réseaux de co-auteurs sur la qualité des publications : plus un chercheur a publié et plus il a eu de co-auteurs différents au cours de sa carrière, plus ses articles sont de bonne qualité.

iii. Introduire la spécialisation des chercheurs, mesurée avec les parts de leur production dans les différents champs de recherche, n'affecte pas les résultats ci-dessus. Néanmoins, on observe certaines disparités disciplinaires et les économistes exerçants en France spécialisés en micro-économie, économie du travail et/ou économie démographique ont publié relativement plus d'articles, de meilleure qualité, et sont plus cités que la moyenne.

iv. La variance des indices de citation est principalement expliquée par les scores de publication, ce qui nous permet de conclure que les pondérations de la qualité des journaux et les citations mesurent globalement la même productivité de la recherche. La relation est naturellement positive entre les deux : plus un chercheur a publié et plus la qualité de ses articles est élevée, plus il est cité.

v. Il y a des effets de réseaux importants dans la diffusion de la connaissance scientifique : à niveau de publications donné (quantité et qualité), les chercheurs ayant travaillé avec des équipes de co-auteurs en moyenne plus grande et qui ont un réseau de stock de co-auteurs différents plus important sont davantage cités.

   Notre interprétation de ce résultat est la suivante : il y a une décomposition entre un effet direct et un effet indirect. Le premier a lieu dans la mesure où les différents co-auteurs d'un même article le présentent à différentes conférences, à différents séminaires et lors de discussions informelles. Le second est à l'œuvre lorsqu'un chercheur parle de ses nouveaux articles aux co-auteurs de ses articles précédents.

Deuxième partie

Après une première partie plutôt descriptive, la deuxième partie de cette thèse cherche à établir des relations causales en économie de la science avec des applications en économie géographique (chapitre 3) et en économie du travail (chapitre 4). En effet, le chapitre 3 s'intéresse aux déterminants individuels et locaux de la productivité des chercheurs et le chapitre 4 s'intéresse aux différentiels de promotion entre hommes et femmes sur le marché du travail académique.

Chapitre 3

Cette étude reprend les méthodologies proposées récemment par l'économie géographique empirique afin de répondre à des questionnements en économie de la science. Nous cherchons à évaluer dans quelle mesure les caractéristiques des universités où sont localisés les chercheurs affectent la productivité de ceux-ci. Même si la plupart des chercheurs ont une opinion personnelle de ce qu'est un bon centre de recherche, il y a étonnement très peu d'études essayant de quantifier cela précisément alors que les implications de politique publique sont importantes. Ce chapitre est donc un des premiers articles du genre.

Nous étudions donc l'impact sur les publications des chercheurs à la fois de leurs caractéristiques individuelles (âge, sexe, statut, etc.) mais aussi d'un nombre important de caractéristiques locales, c'est-à-dire du centre de recherche dans lesquels ils travaillent. Sont considérés la taille du centre, mais aussi sa diversité thématique, sa proximité à d'autres centres de recherche, son hétérogénéité en termes de production de ses chercheurs, sa charge d'enseignement appréhendée via le nombre d'étudiants par chercheur, son nombre de stars (chercheurs ayant une production exceptionnellement élevée) et son ouverture en termes de co-auteurs localisés à l'étranger. Un autre groupe de variables concerne les externalités dites de localisation, relatives aux caractéristiques locales du domaine de recherche dans lequel un chercheur écrit un article. Typiquement, nous étudions dans quelle mesure une forte spécialisation de l'université améliore ou pas la productivité des chercheurs du domaine en question. Enfin, un troisième groupe de variables concerne l'évaluation des externalités émanant de la composition des chercheurs de l'université en termes de statuts (maîtres de conférence, professeurs, Cnrs,...), de sexe et d'âge. Utiliser des données individuelles comme nous le faisons ici nous permet de contrôler, aux côtés des caractéristiques de l'université, des caractéristiques des individus eux mêmes, à nouveau sexe, âge, statut, diversité et connexion avec des co-auteurs à l'étranger.

L'impact des caractéristiques des universités comme des individus est étudié sur leur probabilité de publier, leur nombre de publications et sur la qualité moyenne de celles-ci, appréhendée selon deux indicateurs différents prenant plus ou moins en compte la qualité des revues.

Finalement, ce type d'approche soulevant un certain nombre de problèmes économétriques (variables éventuellement manquantes, tri spatial des chercheurs, causalité inverse liée aux choix endogènes de leur localisation), les stratégies économétriques les plus récentes proposées par l'économie géographique (effets fixes de localisation et individuel, instrumentation) sont mobilisées afin de les circonvenir au mieux.

Les grandes tendances qui se dégagent de notre étude sont les suivantes :

i. Ce sont à la fois ses capacités individuelles et le lieu où il les exerce qui rendent un chercheur productif. En effet, les caractéristiques observables et inobservables des universités où sont localisés les chercheurs ont un pouvoir explicatif au moins équivalent à la moitié du pouvoir explicatif des caractéristiques individuelles des chercheurs.

   Contrairement aux autres papiers de la littérature récente en la matière, nous concluons donc à la présence d'effets de pairs dans la recherche.

ii. Les stratégies individuelles de publication consistant à publier dans différents sous-champs de l'économie, à avoir des co-auteurs localisés à l'étranger et à travailler avec de grandes équipes de co-auteurs augmente la qualité moyenne des publications. Nous interprétons ce dernier effet comme des rendements d'échelle croissants au niveau de l'équipe de recherche. Nous mettons également en évidence des rendements croissants au niveau individuel dans la mesure où le nombre de publications d'un chercheur a un impact positif fort sur la qualité moyenne de ses publications.

iii. Les bons centres de recherche sont caractérisés à la fois par le fait qu'ils attirent des bons chercheurs dans la mesure où nous mettons en évidence un léger tri spatial (les chercheurs les plus productifs sont en moyenne dans des centres de recherche qui génèrent plus d'externalités), mais surtout parce qu'ils génèrent des externalités.

iv. L'homogénéité des chercheurs et la diversité des domaines de recherche ont un effet significativement positif et le pouvoir explicatif le plus important des externalités locales sur toutes les dimensions de la production des chercheurs.

v. La présence de stars, la taille des universités et le fait d'être connecté à des co-auteurs à l'étranger ont un effet positif significatif sur toutes les dimensions de la production des chercheurs mais leur pouvoir explicatif est plus faible.

vi. La proximité spatiale d'autres université a un effet très faible, souvent non-significatif.

vii. La spécialisation d'une université a un rôle significativement positif et important sur la productivité des chercheurs en termes de nombre de publications.

   La spécialisation n'a pas d'effet direct sur la qualité moyenne mais un effet indirect positif sur celle-ci puisqu'elle favorise le nombre de publications et que celui-ci a un effet positif sur la qualité (via les rendements croissants évoqués précédemment).

viii. Parmi les variables de composition, la part des femmes et l'âge moyen ont un impact positif sur la production des chercheurs mais leur pourvoir explicatif est faible. La composition en termes de statuts a un rôle légèrement plus important.

ix. Ces résultats sont robustes à un certain nombre de variantes quant à la taille des universités retenues (supérieure à 4 vs 9 chercheurs), la période d'observation des publications (moyenne mobile sur 3 ans vs décalage de deux ans sans moyenne mobile), le niveau géographique (université vs zone d'emploi), l'endogénéité potentielle de la taille, de la diversité, de l'hétérogénéité et de la part des stars dans le centre de recherche et le nombre d'étudiants moyens par chercheur, qui, contrairement à une intuition possible, n'a pas d'effet sur les publications des chercheurs.

x. Parmi les variables individuelles, les femmes produisent moins que les hommes sans que l'on puisse tester ici l'origine de ce résultat. L'âge a un impact négatif convexe qui disparaît pour devenir positif quand l'effet cohorte est contrôlé gràce aux effets fixes individuels. Les différents statuts ont des niveaux de production conformes à l'intuition, les plus productifs étant les directeurs d'Etudes Ehess et les ingénieurs des ponts et chaussées, suivi des directeurs de recherche Cnrs section 37, des Professeurs des universités de la section 05 du Cnu et ainsi de suite.

Chapitre 4

Ce chapitre étudie une logique inverse au chapitre précédent. En effet, on ne s'intéresse plus à l'impact des caractéristiques individuelles (et donc du statut universitaire) et des caractéristiques des universités dans lesquelles les chercheurs travaillent sur leur productivité mais on regarde dans quelle mesure la production scientifique passée des (enseignants)-chercheurs, qui est une mesure de leur productivité moyenne individuelle, et leurs caractéristiques démographiques affectent leur probabilité d'être promu et/ou de travailler dans un bon centre de recherche. Plus particulièrement, nous nous intéressons ici au différentiel de poste atteint entre les hommes et les femmes, à l'instar de ce qui est fait en économie du travail. Cependant, la littérature sur les différentiels entre hommes et femmes sur le marché du travail se concentre principalement sur les écarts de salaire dans la mesure où les différentiels de promotion dépendent davantage de caractéristiques individuelles inobservables. Cette étude est d'autant plus intéressante parce qu'il y a très peu de données pour lesquelles il est possible d'observer la productivité individuelle des travailleurs. Utiliser des données de recherche académique nous permet de mesurer la productivité individuelle des chercheurs avec leurs publications. Ainsi nous passons outre cette limite statistique qui existe sur la plupart des marchés du travail alors que les différentiels de promotion entre hommes et femmes sont toujours importants aujourd'hui sur la plupart de ces marchés.

Parmi les économistes universitaires (et assimilés) exerçants en France, il y a principalement deux types de statuts universitaires : ceux que nous appelons “rang B” correspondent aux emplois dits parfois “junior”, c'est-à-dire les Maître de conférences (université) et les Chargés de recherche (Cnrs, Inra) et ceux que nous appelons “rang A” correspondent aux chercheurs parfois appelés “seniors” qui ont été promus au grade de Professeurs des universités ou de Directeurs de recherche (Cnrs, Inra). Sur l'ensemble de la période étudiée (1990-2008), 43% des hommes sont rang A contre seulement 18% des femmes. Une partie importante de cet écart s'explique par le fait que les hommes sont en moyenne plus âgés (alors qu'être promu requiert du temps) et qu'ils ont des stocks de publications en moyenne plus importants. Néanmoins, à âge et publications données, les femmes sont toujours moins souvent rang A que les hommes.

Nous avons alors testé si la partie non expliquée de cet écart pouvait être due à de la discrimination contre les femmes, à un coût de promotion plus élevé pour les femmes, ou à des préférences différentes entre salaire et prestige du centre de recherche, entre les hommes et les femmes. Pour ce faire, nous avons d'abord utilisé les particularités du système universitaire et plus particulièrement du système universitaire français. Premièrement, dans le système universitaire, une promotion n'est pas automatiquement associée à des contraintes ou à des obligations supplémentaires, contrairement à la plupart des emplois du secteur privé. Dans les faits, les chercheurs seniors ont, en moyenne, davantage de travail administratif que les chercheurs juniors, mais ce n'est absolument pas statutaire. Ainsi, même si les femmes consacrent davantage de temps que les hommes à la vie familiale par exemple, il n'y a pas de raison qu'elles ne cherchent pas à être promues pour autant. Dans le système français, il y a aussi une échelle salariale nationale et la présence d'un concours national pour passer Professeur des universités, concours qui nécessitera, si obtenu, de changer d'université de rattachement et donc probablement de ville alors que les chercheurs au Cnrs ou à l'Inra n'ont pas cette obligation de mobilité en cas de promotion. Cette existence de deux catégories de chercheurs publics, des chercheurs purs au Cnrs ou à l'Inra et des enseignants-chercheurs à l'université, dont l'une doit muter en cas de promotion, est cruciale dans notre analyse.

Discrimination : Pour tester une éventuelle présence de discrimination contre les femmes, nous avons comparé l'effet du sexe sur la probabilité d'être promu(e) et sur la probabilité de travailler dans un bon centre de recherche. En effet, comme indiqué par Lazear et Rosen (1990), il serait difficilement compréhensible que les femmes soient discriminées dans une dimension (promotion) mais pas dans une autre (qualité de l'institution d'accueil).

Coût de promotion plus important : Une autre explication possible de l'important différentiel de poste atteint entre les hommes et les femmes seraient la présence d'un coût de promotion plus élevé pour les femmes. Pour tester cette hypothèse, nous utilisons le fait qu'il y a deux types de chercheurs publics en économie en France : les universitaires de la section 05 du Cnu qui doivent changer d'université et donc probablement de ville si ils obtiennent le statut de Professeur des universités par l'Agrégation du supérieur et les chercheurs purs (Cnrs et Inra) qui peuvent rester dans leur centre de recherche d'accueil en cas de promotion. Si les coûts de promotion imposés par la mobilité des universitaires étaient plus élevés pour les femmes, alors elles devraient être moins souvent Professeurs des universités que Directrices de recherche au Cnrs ou à l'Inra, toutes choses égales par ailleurs. Cela est d'autant plus possible que l'âge moyen des candidats à l'Agrégation du supérieur se situe entre 30 et 40 ans, soit à un moment du cycle de vie où les contraintes familiales peuvent être importantes.

Préférences différentes entre salaire et prestige : La dernier explication que nous pouvons tester pour expliquer le différentiel de poste atteint entre les hommes et les femmes pourrait être la suivante : ayant encore majoritairement un salaire d'appoint dans le ménage, les femmes universitaires préfèrent rester Maître de conférences quand elles sont dans des institutions prestigieuses plutôt que de prendre le risque probable d'être mutées dans de moins bons centres de recherche en cas de promotion. Là encore, nous pouvons utiliser le fait que les chercheurs purs (Cnrs ou Inra) ne doivent pas changer d'institution en cas de promotion pour tester cette hypothèse. En effet, si les femmes sont moins prêtes que les hommes à renoncer au prestige de leur institution pour augmenter leurs salaires, alors celles qui sont universitaires dans une bonne institution candidateront moins à une promotion, sachant qu'elles risqueraient de se retrouver dans un moins bon centre de recherche en cas de succès. Au contraire, comme aucune mobilité n'est exigée en cas de promotion pour les chercheurs purs (Cnrs, Inra), les femmes doivent y chercher davantage une promotion.
Les résultats de ces différents tests sont les suivants :

i. Dans la mesure où, d'une part, même si les femmes ont une probabilité plus faible que les hommes, à caractéristiques individuelles observables équivalentes, d'être Professeur des universités ou Directrice de recherche, elles ont une probabilité plus grande de travailler dans une institution académique prestigieuse, et que, d'autre part, les femmes ont des rendements des publications supérieurs aux hommes à la promotion, nous concluons à l'absence de discrimination contre les femmes dans le sytème universitaire français.

ii. Nous ne trouvons pas que l'impact négatif d'être une femme sur la probabilité de promotion soit plus important pour les universitaires (à qui une mobilité est imposée en cas de promotion par le concours de l'Agrégation du supérieur) que pour les chercheurs purs (Cnrs et Inra) qui n'ont pas à bouger en cas de promotion. L'hypothèse selon laquelle les coûts de promotion seraient plus élevés pour les femmes est donc rejetée.

iii. Dans la mesure où les femmes ont une probabilité plus importante d'être rang A dans un moins bon centre de recherche au Cnrs ou à l'Inra que les femmes universitaires, nous en concluons qu'elles ne font pas un arbitrage différent des hommes entre salaire et prestige de l'institution dans laquelle elles travaillent.

iv. Si le différentiel de poste atteint qui n'est pas expliqué par les caractéristiques démographique (âge) ou de performances (publications) reste important entre les hommes et les femmes, cela pose alors la question de savoir si les femmes s'autocensurent en ne cherchant pas à être promues ou en faisant moins d'efforts que les hommes durant les processus de promotion.

Troisième partie

Dans un article relativement récent et malgré cela déjà très influent (très cité), Santos Silva et Tenreyro (2006) ont argumenté contre l'estimation des modèles log-linéarisés par la méthode des moindres carrés ordinaires, jugeant que le processus de log-linéarisation biaisait l'estimation des paramètres d'intérêt en cas d'hétéroscédasticité en niveau. Cette analyse repose sur l'inégalité de Jensen (l'espérance du logarithme d'une variable aléatoire n'est pas égale au logarithme de son espérance) et peut être exposée de la manière suivante : si la variance des résidus du modèle en niveau est fonction de variables explicatives (hétéroscédasticité), log-linéariser le modèle va rendre les nouveaux résidus fonction de ces variables explicatives dans la mesure où le logarithme d'une variable aléatoire ne dépend pas uniquement de son espérance mais aussi d'autres moments d'ordre supérieur dont sa variance. Cela viole alors la condition sous laquelle les moindres carrés ordinaires sont consistants : l'indépendance des résidus aux variables explicatives.

Santos Silva et Tenreyro (2006) préconisent alors d'utiliser un estimateur du modèle directement en niveau et précisent que c'est l'estimateur du pseudo-maximum de vraisemblance à partir d'une loi de Poisson qui a leur préférence, dans la mesure où celui-ci devrait être plus efficace, ce qu'ils illustrent par des simulations Monte Carlo. Introduit initialement par Gourieroux, Monfort et Trognon (1984a,b), cet estimateur, comme les autres estimateurs de pseudo- (ou quasi-) maximum de vraisemblance on été utilisés dans un premier temps pour des données de comptage.

Chapitre 5

Dans cette partie, je commence d'abord dans le chapitre 5 par relativiser ce résultat en montrant qu'il repose sur l'hypothèse forte selon laquelle les résidus d'un modèle doivent être spécifiés en niveau alors que cette hypothèse, bien que possible, n'a ni moins ni davantage lieu d'être que son alter ego consistant à spécifier des résidus au modèle log-linéarisé. En effet, si le "vrai" modèle inclue des résidus en niveau (avec hétéroscédasticité), alors l'estimation du modèle log-linéaire conduit à l'obtention de coefficients biaisés ; mais si le "vrai" modèle inclue des résidus additifs au log-linéaire (ici aussi avec hétéroscédasticité), alors c'est l'estimation par un estimateur non-linéaire du modèle en niveau qui va conduire à l'obtention de coefficients biaisés. Ce résultat de réciprocité de la proposition de Santos

Silva et Tenreyro (2006) n'avait jusqu'alors jamais été discuté, il est confirmé ici par des simulations Monte Carlo. Dans un cadre appliqué, sans connaissance préalable sur la nature du “vrai” modèle générateur des données, il pourrait alors être intéressant de pouvoir tester de manière discriminatoire un modèle contre un autre, par exemple le modèle log-linéaire contre un modèle non-linéaire en niveau dans la mesure où ils peuvent conduire à des estimations différentes.

Pour ce faire, une possibilité réside dans l'utilisation d'un test de spécification avec hypothèses alternatives non-emboitées. Le test P_ϵ (test P étendu) de MacKinnon, White et Davidson (1983), extension des tests J et P de Davidson et MacKinnon (1981), en particulier au cas où les deux modèles non-emboités impliquent des transformations différentes de la variable dépendante, cas qui nous intéresse ici, pourrait présenter une solution intéressante. Malheureusement, ce test ne peut être utilisé qu'en présence de résidus identiquement distribués suivant une loi Normale et ne peut donc servir qu'à discriminer entre une estimation par moindres carrés ordinaires du modèle log-linéarisé et une estimation par moindres carrés non-linéaires du modèle en niveau. Des simulations Monte Carlo confirment en effet que le test permet de discriminer entre une estimation du modèle log-linéaire par moindres carrés ordinaires et une estimation du modèle non-linéaire en niveau lorsque la variance des résidus en niveau est constante, c'est-à-dire lorsque l'estimateur par moindres carrés non-linéaires est théoriquement optimal. Dans les autres cas étudiés ici, dans lesquels on introduit des formes particulières d'hétéroscédasticité pour lesquelles des estimateurs de pseudo-maximum de vraisemblance à partir d'une loi de Poisson (variance conditionnelle proportionnelle à l'espérance) ou à partir d'une loi gamma (variance conditionnelle proportionnelle à l'espérance au carré) sont optimaux, le test est majoritairement non-discriminant et parfois trompeur. Cela signifie qu'il conduit à rejeter le bon model dans ce dernier cas et qu'il ne permet pas de trancher entre les deux hypothèses alternatives non-emboitées dans le premier cas, soit parce que chaque spécification rejette l'autre quand elle est vraie sous l'hypothèse nulle, soit au contraire parce qu'elles s'acceptent mutuellement. Ce résultat constitue la limite principale de ce chapitre dans la mesure où on aimerait aussi pouvoir tester l'utilisation d'un estimateur de pseudo-maximum de vraisemblance à partir d'une loi de Poisson (sous l'hypothèse sous-jacente qu'il est optimal) dans la lignée des travaux de Santos Silva et Tenreyro (2006).

Une application aux équations de gravité confirme également ce résultat dans la mesure où tous les tests effectués sont non-conclusifs, c'est-à-dire qu'ils ne permettent pas de discriminer entre les différentes spécifications possibles.

Chapitre 6

Dans la littérature du commerce international, et plus précisément des équations de gravité, il existe un paradoxe important appelé “paradoxe de la distance” ou “paradoxe de la mondialisation manquante” (missing globalisation puzzle). Celui ci peut s'énoncer de la manière suivante : alors que l'on observe une baisse continue des coûts de transport au cours de la seconde moitié du 20ème siècle, le paramètre associé à la distance dans les équations de gravité augmente sur la même période (en valeur absolue). Il est d'abord utile de noter que ce paramètre est une élasticité et que, pour qu'elle diminue, il faut que l'effet marginal de la distance sur le commerce international diminue, pas son effet total. Notons aussi que cette élasticité du commerce à la distance est elle même le produit de deux élasticités : l'élasticité du commerce aux coûts de commerce et l'élasticité des coûts de commerce à la distance. Même si la seconde diminue, la première peut augmenter suffisamment pour que le produit reste stable au cours du temps.

Cela dit, on peut considérer qu'il est quand même assez surprenant de voir le paramètre associé à la distance dans les équations de gravité non pas stagner (ce qui serait compréhensible au vu de ce qui est énoncé ci-dessus) mais augmenter au cours du temps. Compte tenu de ce qui a été exposé précédemment (chapitre 5), il est alors important de noter que ce résultat est obtenu à l'aide d'estimations par moindres carrés ordinaires du modèle de gravité log-linéarisé. Il n'est pas dit que ces estimations soient fausses mais elles pourraient tout aussi bien l'être et il serait alors intéressant de savoir ce qu'on obtient en utilisant un autre estimateur.

Dans ce chapitre, nous montrons alors que l'utilisation d'un estimateur de pseudo-maximum de vraisemblance à partir d'une loi de Poisson pour un modèle gravitaire en niveau conduit à l'obtention d'une élasticité du commerce à la distance constante dans le temps, résultat qui parait plus conforme à l'intuition. Nous montrons également que l'écart grandissant entre les estimations par moindres carrés ordinaires du modèle log-linéarisé et par pseudo-maximum de vraisemblance à partir d'une loi de Poisson du modèle en niveau est dû à l'hétérogénéité croissante des flux de commerce.

Chapitre 7

Enfin, certains auteurs ont utilisé un estimateur de pseudo-maximum de vraisemblance à partir d'une loi binomiale négative, par extension à l'estimateur du pseudo-maximum de vraisemblance à partir d'une loi de Poisson proposé par Santos Silva et Tenreyro (2006), pour estimer des équations de gravité et d'autres modèles en niveau. Cet estimateur permet l'estimation jointe des paramètres d'intérêt du modèle et d'un paramètre de sur-dispersion (variance conditionnelle plus que proportionnelle à l'espérance conditionnelle). Il a l'avantage d'inclure les estimateurs de pseudo-maximum de vraisemblance à partir d'une loi de Poisson et à partir d'une loi gamma, comme des cas extrêmes du modèle basé sur une loi binomiale négative quand le paramètre de sur-dispersion prend respectivement la valeur *zéro* ou *plus infini*. Il peut donc être considéré à ce titre comme plus général.

Cependant, nous montrons dans ce chapitre que les estimateurs (en une ou deux étapes) de pseudo-maximum de vraisemblance à partir d'une loi binomiale négative, qui ont été utilisé dans la littérature appliquée aux équations de gravité ou à d'autres modélisations, sont fonctions de l'échelle/unité de mesure de la variable dépendante. Cela veut dire que les paramètres estimés du modèle ne seront pas les mêmes si la variable dépendante est exprimée en milliers d'euros ou en millions de dollars. Si la variable dépendante est une donnée de comptage, ce pour quoi sont utilisés ces estimateurs à l'origine, cela n'est pas gênant puisqu'une variable de comptage n'a pas d'unité autre que le compte. Mais si la variable dépendante est en données monétaires par exemple, comme un salaire ou un flux de commerce, alors il n'est pas souhaitable pour un estimateur d'avoir cette propriété. Il n'y a en effet pas de raison pour que l'estimation de l'élasticité du commerce à la distance soit fonction du choix de la monnaie utilisée pour enregistrer les données de commerce. Ce problème est assez important dans la mesure où certains paramètres du modèle gravitaire peuvent aller du simple au double lorsque l'on estime une équation de gravité avec des estimateurs de pseudo-maximum de vraisemblance à partir d'une loi de Poisson ou d'une loi gamma, cas extrêmes de la modélisation à partir d'une loi binomiale négative.

Cette propriété négative de dépendance à l'échelle/unité de mesure de la variable dépendante de l'estimateur de pseudo-maximum de vraisemblance à partir d'une loi binomiale négative disparait lorsque l'on utilise un nouvel estimateur de pseudo-maximum de vraisemblance à partir d'une loi binomiale négative en deux étapes que nous proposons. Cette solution est permise grâce à l'utilisation de l'hypothèse de variance de modèle linéaire généralisé (GLM variance assumption, Wooldridge (1999)).

Première partie

Measurement of Research Outputs and Determinants of Individual Records

Chapitre 1

Un panorama de la recherche française en économie en comparant les approche Google Scholar et Econlit [1]

Nous utilisons les citations Google Scholar 2010 des économistes exerçants en France en 2008 afin de dessiner un panorama de la recherche en économie en France et de tester l'utilisation d'un support plus large qu'Econlit pour l'évaluation de la production de

1. Ce chapitre est une version très légèrement actualisée du document de travail du GREQAM n° 2011-56 coécrit avec Pierre-Philippe Combes. Il trouve son origine dans un rapport intitulé "Comparaison des mesures Econlit et Google Scholar de la production de recherche en économie en France en 2008" que nous avons produit pour la la Direction Générale de la Recherche et de l'Innovation (DGRI) du Ministère de l'Enseignement Supérieur et de la Recherche. Nous remercions vivement la DGRI pour son soutien financier et Marc Ivaldi pour ses précieux conseils. Nous remercions également Philippe Donnay et Charles Laitong pour leur excellent travail d'assistance de recherche. Les opinions émises dans cet article ne représentent que celles des auteurs et non celles de la DGRI ou du Ministère de l'Enseignement Supérieur et de la Recherche. L'article correspondant est à paraître dans la Revue d'Économie Politique.

recherche. Nous comparons les indicateurs de citations tels que le nombre de citations divisés par le nombre d'auteurs, l'indice H ou l'indice G calculés au niveau individuel mais aussi des centres de recherche et universités avec des indices de publications calculés via la base de données EconLit et prenant plus ou moins en compte la qualité des journaux. Ces comparaisons sont menées sur différentes périodes de temps.

La hiérarchie des institutions calculée avec l'approche Google Scholar est relativement proche de celle observée lorsque l'on utilise des scores de publications. Néanmoins, on observe également certaines variations spectaculaires pour quelques institutions, positives principalement lorsque l'économie n'est pas le cœur de métier de ces institutions.

1.1. Introduction

Les classements des centres de recherche et universités passent la plupart du temps par une inférence indirecte de la qualité de leurs publications par la qualité du journal dans lequel elles sont publiées. Le principal reproche fait à ce type d'approche est la relativement grande variabilité de qualité qu'il peut rester entre les différents articles d'un même journal. Il est alors souvent proposé d'utiliser directement le nombre de citations reçues par chaque publication. Cela suppose tout d'abord que ce nombre de citations est un bon indicateur de la qualité de la publication, ce qui est généralement admis. Il ne faut pas oublier que cette hypothèse n'est pas complètement triviale, la littérature soulignant par exemple les fortes différences de pratique de citations (nombre de références bibliographiques par article par exemple) entre domaines, et naturellement les différences de nombre de chercheurs de chaque domaine qui affectent directement le nombre de citations susceptibles d'être reçues. La nécessaire pondération des citations reçues par la qualité de la publication citant (comme cela est fait pour évaluer l'impact des journaux par exemple) a priori également nécessaire est souvent difficile à implémenter. Finalement, le nombre de citations d'une publication est très affecté par son cycle de vie, ce qui rend crucial le choix de la fenêtre temporelle pendant laquelle les citations sont recueillies, en fonction de l'âge de la publication. Malgré ces limites, utiliser les citations de chaque publication est souvent considéré comme souhaitable mais rarement effectué du fait du nombre très limité de bases de données le permettant.

La principale source de citations, le Journal of Citation Reports (JCR) de Thomson-Reuters, fournit les citations reçues par les articles publiés dans 304 journaux en économie, alors qu'on en recense plus de 1200 dans la base de données Econlit par exemple (qui regroupe la plupart des revues de recherche en économie). Il est important de noter que la limitation est double : d'une part, une publication n'est citée que si elle appartient à un de ces 304 journaux, d'autre part, les citations ne sont recueillies que dans ces mêmes 304 journaux. Malgré cela, et en ce qui concerne les économistes français, Combes et Linnemer (2003a) comparent pour la France différentes approches qui utilisent les citations JCR afin de hiérarchiser l'impact des publications en économie des centres de recherche et universités françaises en 1998. L'objet du présent article est de déterminer si l'utilisation de Google Scholar (GS) pourrait constituer une alternative intéressante à ces deux stratégies, inférence indirecte de la qualité par celle du journal ou citations du JCR.

GS est un outil qui permet de considérablement dépasser les deux limites que présentent le JCR. GS recense les publications, de tout type, présentes sur des pages Internet académiques et calcule les citations qu'elles reçoivent sur des supports eux-mêmes présents sur Internet, avec la simple condition qu'ils soient également de nature académique. [2] Ainsi, est

2. On peut lire sur le site Google Scholar "[Il sagit d'] articles revus par des comités de lecture, thèses, livres, résumés analytiques et articles. Ces travaux peuvent provenir de sources telles que des éditeurs scien-

élargi à la fois le type de publications susceptibles de recevoir des citations, ce qui prend notamment en compte les supports potentiellement importants pour les économistes que constituent les ouvrages ou les documents de travail, mais élargit aussi considérablement le nombre de supports citant. De plus, le domaine d'étude est a priori moins restreint que lorsqu'on le définit ex-ante en le réduisant aux journaux référencées par Econlit ou encore plus par le JCR. Si un économiste a une publication importante dans un journal en mathématiques non référencé par Econlit ou le JCR, celle-ci est en général considérée via GS.

Les limites de GS sont principalement dues au fait qu'il s'agit d'un outil récent, et donc sans aucun doute encore en train de s'améliorer mais sujet à imprécisions, ainsi qu'au fait que GS fonctionne fondamentalement selon la même philosophie que Google, à savoir une recherche de proximité de mots entre diverses entrées, ce qui pose dès le départ la question du seuil à partir duquel on considère que deux groupes de mots sont identiques ou pas. Ainsi, un certain bruit dans la définition de ce qu'est une publication est tout d'abord présent, nous parlerons d'ailleurs souvent d'"entrée" GS plutôt que de publication. Par exemple, différents chapitres d'un ouvrage peuvent constituer autant d'entrées différentes, il en va de même pour un même article publié dans différentes séries de documents de travail, éventuellement à des dates différentes, ou sous des titres légèrement différents, même si un effort certain de regroupement, justement grâce à des algorithmes de proximité, est effectué par GS lui même. Ensuite, on retrouve au niveau des supports citant des sources de bruit de même type. Si deux versions d'un même article considérées comme deux entrées différentes citent un même travail, celui-ci reçoit deux citations. Finalement, les variables relatives à

tifiques, des sociétés savantes, des référentiels de pré-publication, des universités et d'autres organisations de recherche."

chaque entrée GS sont elles mêmes de qualité nettement moindre que dans Econlit ou JCR. Le nom des co-auteurs est moins précis (absence plus fréquente du prénom complet par exemple), leur nombre est plus difficile à calculer (en partie du fait des imprécisions sur les noms et prénoms), la date de publication n'est pas toujours disponible, ou entachée d'erreur, etc. Cependant, et comme toujours, la seule question importante pour le statisticien est de savoir si les erreurs de mesures sont corrélées ou pas avec le phénomène qu'il tente d'évaluer. Lorsque les erreurs sont distribuées aléatoirement, elles ne gênent en général pas l'analyse.

Malgré ces limites qu'il est crucial de garder en tête, nous proposons ici une étude prospective ayant pour objet de présenter un panorama de la recherche française en économie et d'étudier les propriétés d'un certain nombre d'indicateurs d'impact des centres de recherche et universités françaises en 2008 en économie. Ces indicateurs sont fondés sur les citations GS que reçoivent en janvier 2010 les entrées antérieures à 2008 dans les sujets "Business, Administration, Finance, and Economics" et "Social Sciences, Arts, and Humanities" de leurs membres. Nous calculons, pour tout période de temps T cinq indicateurs de citation GS : le nombre total de citations ($Ct(T)$), le nombre de citations par entrée GS ($Ce(T)$), le nombre de citations par entrée GS ayant reçu au moins une citation ($Cp(T)$), l'indice H ($H(T)$), et l'indice G ($G(T)$), tous ces indices étant dupliqués selon que l'on prend en compte ou pas le nombre de co-auteurs. Cinq périodes de temps différentes sont considérées : toutes les années (T=All), les cinq dernières années (2004-2008, T=5 ans), en décomptant dans le temps (T=Dégressif), par année (d'existence de la publication, T=Annuel), et par année de carrière (du chercheur, T=Carrière). Les résultats obtenus ne sont pas fondamentalement différents entre certaines de ces variantes. Ainsi, avons-nous choisi de nous concentrer dans les sections de classements sur le nombre de citations totales et l'indice G, en volume et par chercheur, et sur la période T=Dégressif. Nous présentons quelques variantes particulièrement intéressantes dans des sections spécifiques et un grand nombre de résultats complémentaires sont donnés dans Bosquet et Combes (2011). Les résultats obtenus sont systématiquement comparés à ceux utilisant une approche plus traditionnelle basée sur Econlit et la pondération de la qualité moyenne des journaux, de façon modérée (Clm) ou forte (Clh), proposée par Combes et Linnemer (2010).

Les grandes tendances qui se dégagent de notre étude sont les suivantes. Tout d'abord, GS constitue un support de publications entre deux et quatre fois plus large qu'Econlit. Les chercheurs publiant selon Econlit ont plus d'entrées GS que de publications Econlit et la part des non-publiant est plus faible selon GS. Seuls 6.2% des chercheurs n'ont pas d'entrée GS et 15% n'en ont pas ayant au moins une citation, alors que 27.1% n'ont pas de publication Econlit (respectivement, 22.4%, 34.3% et 45.3% sur 2004-2008). L'ensemble des chercheurs localisés en France ont ensemble quasiment 265000 citations GS en janvier 2010, 47000 pour leurs entrées de la période 2004-2008. Chaque entrée a en moyenne presque 7 citations, le double pour celles en ayant au moins une. L'ensemble des entrées GS de la France entraîne un indice H collectif d'environ 180, soit 180 entrées ayant au moins 180 citations et environ 500 entrées ayant en moyenne 500 citations (indice G).

Prendre en compte ou pas le nombre de co-auteurs ou ramener les citations par entrée ayant ou pas au moins une citation n'a qu'une influence très marginale sur les classements des centres et universités. Il en va de même pour le fait d'utiliser les indices H ou G, ou de modifier la prise en compte du temps. Il est important de noter que les approches GS, quelles qu'elles soient, conduisent à une hiérarchie des universités ou centres très proche de celle obtenue via l'approche Clm qui se base sur Econlit. Ainsi, utiliser la qualité des journaux comme prédicteur de la qualité des publications, en tous les cas du nombre de citations GS qu'elles reçoivent, semble constituer une stratégie pertinente, au moins au niveau agrégé des centres et universités.

Notons néanmoins les quelques résultats particuliers suivants. Les indicateurs de citations GS sont moins sélectifs que ne le sont les indicateurs Clh, et se rapprochent plus des indicateurs Clm. Les corrélations entre classements GS et Econlit sont légèrement moins élevées en ce qui concerne la hiérarchie des institutions selon leurs scores par chercheur. La production des chercheurs pour les indices GS, quels qu'ils soient, suit une courbe en cloche au cours de leur cycle de vie, comme on l'obtient pour les indices Clm et Clh, les chercheurs ayant un pic de productivité généralement observé entre 40 et 50 ans. Les quatre ou cinq statuts les plus productifs selon les indices GS restent identiques à ceux identifiés au moyen de Clm et Clh (Ingénieurs Ponts et Chaussées, Directeur d'Etudes Ehess, Administrateurs Insee et Directeurs de recherche Cnrs). En revanche, progressent selon les mesures GS les statuts qui ne sont pas au coeur de notre analyse (hors section 5 Cnu ou section 37 Cnrs, notamment les "assimilés" professeurs ou chargés de recherche, en poste dans des grandes écoles, parfois de commerce, ou dans des administrations), l'élargissement du support de publication en étant la cause probable.

En ce qui concerne le classement des universités et centres, il découle de ces résultats que d'une part, pour une majorité de centres ou universités, la variation de classements par rapport à ceux obtenus selon l'indice Econlit Clm est faible. D'autre part, pour quelques autres, elles est relativement forte, et il s'avère que ces unités semblent être celles dont le cœur de l'activité n'est pas l'économie, ou, ce qui est lié, dont les membres ne sont pas au cœur de notre champ, à savoir les sections 5 du Cnu et 37 du Cnrs. C'est cette dernière propriété qui peut faire de GS un instrument complémentaire intéressant de l'instrument Econlit bien que susceptible d'être plus entaché d'erreurs de mesure.

1.2. Champ de l'étude et choix méthodologiques

1.2.1. Chercheurs, centres et universités

Nous considérons 104 centres de recherche français en économie (section 5 du Cnu ou section 37 du Cnrs) d'au moins 5 chercheurs. Ces centres sont soit des Unités Mixtes de Recherche (Université, Grandes Ecoles et Cnrs ou Inra), soit l'ensemble des enseignants-chercheurs d'une université n'appartenant justement pas à de telles Umr. Ces centres s'agrègent en 75 'universités' (qui peut être en fait une Grande Ecole). En moyenne, un centre a 27 chercheurs et une université en rassemble 38, 2832 chercheurs équivalent temps plein étant pris en compte dans l'étude. Le vocable chercheur utilisé ici correspond soit à des enseignants-chercheurs de l'Université ou des Grandes Ecoles, soit à des chercheurs Inra ou Cnrs. Un chercheur peut être affilié à plusieurs centres ou universités, voire être en partie aussi affecté à des universités étrangères, d'où la notion d'équivalent temps (en France). Un système de poids, égalitaire dans la plupart des cas, mais parfois légèrement différent notamment pour les affiliations à l'étranger, donne la clé de répartition du chercheur entre ses différentes affiliations. Les caractéristiques démographiques des chercheurs et leur répartition entre différentes institutions ou status peuvent être trouvées dans Bosquet, Combes et Linnemer (2010). Rappelons simplement ici qu'environ 90% des chercheurs n'ont qu'une seule affiliation, mais que les 10% en ayant plus d'une rassemblent environ

40% de la production, ce qui souligne l'importance des choix de ventilation de celle-ci entre les différentes affiliations. Par ailleurs, la structure par âge des chercheurs est clairement bi-modale, avec un premier pic aux environs de 38 ans et un deuxième autour de 60 ans.

1.2.2. Mesure des stocks de publications via Econlit

Nous commençons par rappeler les caractéristiques des approches Clm et Clh basées sur Econlit proposées par Bosquet et al. (2010) auxquelles nous allons comparer les approches fondées sur GS. Les publications prises alors en compte correspondant à la catégorie "Journal article" de la version de juin 2009 de la base de données Econlit de l'*American Economic Association*, dont nous extrayons les publications jusqu'en 2008 inclus. Cette base considère en 2008 plus de 1200 journaux, pour un total de 556 770 articles publiés entre 1969 et 2008. Pour chaque article, les informations suivantes sont utilisées : date de publication, journal où il a été publié, nombre de pages, nombre de co-auteurs, codes Jel de classification des domaines. Les mesures de stock de publications d'un chercheur dépendent de deux éléments cruciaux, le système de pondération des journaux dans lesquelles les articles sont publiés, W (pour "Weight"), et le système de pondération des articles en fonction de la période où ils sont publiés, T (pour "Time"). Deux autre dimensions sont également prises en compte, le nombre de co-auteurs ainsi que la longueur de l'article. Ces différents éléments sont combinés de la manière suivante.

Soit $A(i)$ l'ensemble des articles (référencés dans Econlit et donc publiés entre 1969 et 2008 ici) du chercheur i. Soit a un article dans $A(i)$. Chaque article est caractérisé par son nombre d'auteurs $n(a)$, son nombre de pages $p(a)$, son année de publication $t(a)$, la longueur moyenne d'un article publié la même année dans le même journal $\overline{p}(a)$ et enfin le poids accordé au journal où il a été publié $w(a)$ appartenant au système de pondération des journaux W. Le stock de publications du chercheur i selon W sur la période T est évalué à l'aide de la formule suivante :

$$y_i\,(T,W) = \sum_{a\in A(i)} T\,(t(a))\,\frac{w(a)}{n(a)}\frac{p(a)}{\overline{p}(a)} \tag{1.1}$$

où $T(.)$ est la fonction qui pondère les années, détaillée plus bas.

Au niveau individuel, la manière de prendre en compte le nombre de co-auteurs fait débat. Tout le monde s'accorde sur le fait que publier seul un article représente une plus forte contribution que de le faire à plusieurs. Supposer, comme nous le faisons ici qu'une publication co-écrite à n auteurs compte n fois moins qu'un article publié seul semble à certains trop extrême. Des auteurs (par exemple Lubrano, Bauwens, Kirman et Protopopescu, 2003) proposent non pas de diviser par $\frac{1}{n}$ mais plutôt par $\frac{1}{\sqrt{n}}$. Cela pose cependant à la fois un problème d'interprétation économique et un problème d'agrégation. Si deux personnes écrivent ensemble un article, $\frac{1}{\sqrt{2}} \approx 0.71$ article est attribué selon cette méthode à chaque auteur. Autrement dit, en joignant deux productions de 0.71 article, on obtient un seul article. Cette hypothèse postule donc implicitement que l'on aurait des rendements d'échelles décroissants dans la production d'articles de recherche. On a alors du mal à comprendre pourquoi autant de gens choisissent de co-publier. De plus, au niveau agrégé d'un centre de recherche par exemple, cette méthode conduit à comptabiliser l'article comme 1.42 article si les deux auteurs sont membres du centre en question alors qu'un seul article est réellement publié par le centre. Enfin, cette comptabilisation donnerait clairement des incitations à ajouter le maximum d'auteurs du centre sur toutes ses publications, ce qui semble d'ailleurs se produire dans certains domaines.

A l'inverse, si l'on pense que les chercheurs travaillent en équipe justement pour bénéficier de complémentarités et de rendements d'échelle croissants, on pourrait vouloir utiliser une fonction puissance supérieure à 1 (en valeur absolue). Par exemple, pourrait être attribué à chaque auteur seulement $\frac{1}{n^{1.5}}$ de l'article. Dans le cas de deux auteurs, chacun obtiendrait une contribution de 0.35, soit 0.7 à deux, les 0.3 “manquant” pour obtenir un article représentant l'impact de l'externalité. Mais si l'on pense que cette externalité doit tout de même être attribuée aux auteurs de la publication, et par suite à leur centre ou université, chacun atteint bien une part de $0.35 + \frac{0.3}{2} = 0.5$, ce qui correspond à notre approche.

Finalement, notons que l'économie se distingue d'autres domaines par le fait que les auteurs ne sont que très rarement hiérarchisés et simplement indiqués par ordre alphabétique. La profession admet donc explicitement que les bénéfices de la publication doivent être équitablement partagés entre auteurs, ce que notre approche choisit également. Avoir des auteurs présentés en ordre non alphabétique soulève des problèmes auxquels il n'existe, autant que nous le sachions, pas de solution faisant consensus. Typiquement s'il est admis que le premier auteur se voit attribuer une fraction de l'article supérieure à l'inverse du nombre d'auteurs, chacun semble avoir un avis personnel sur ce que doit être cette fraction, et cette question se pose pour chaque rang de co-autorat. Le point de vue égalitaire des économistes, quasiment toujours adopté même si quelques exceptions existent, simplifie donc grandement la tâche à ce niveau.

Lorsque tous les articles d'Econlit sont considérés, la répartition est approximativement de 60% des articles écrits seul, 30% écrits à deux et 10% écrits à trois ou plus. Sur les cinq dernières années, le pourcentage d'articles écrits seul diminue pour devenir inférieur à 50%. Par rapport à ces tendances obtenues sur l'ensemble d'Econlit, les chercheurs français publient moins seuls, n'étant désormais plus qu'environ un tiers à le faire.

Prendre en compte la longueur des articles est un choix naturel. L'article median d'Econlit a 15 pages mais 25% des articles en ont plus de 21 et 10% plus de 28. À l'autre extrême, 25% des articles ont moins de 9 pages et 10% moins de 5. La disparité est importante. On peut penser que le nombre d'idées et innovations dans les articles les plus courts est moins important que dans les plus longs. Toutefois, une question de croissance des rendements émerge là aussi : la dernière page d'un très long article a-t-elle la même importance que celle d'une article très court ? Comme pour les co-auteurs, une idée serait de considérer non pas le nombre de pages publiées mais sa racine pour prendre en compte cette décroissance des rendements. Bien que l'argument soit pertinent, le choix de la forme fonctionnelle reste délicat. De plus, un tel choix pose, comme pour les co-auteurs, des problèmes d'agrégation, puisque scinder un article en deux améliorerait le score (même si cette pratique est parfois déjà observée du fait d'un grand nombre de mesures reposant sur le seul nombre de publications).

Le deuxième problème est celui de la comparaison du nombre de pages entre les différents journaux. D'une part, la typographie et la taille des pages diffèrent entre journaux. Par exemple, une page dans l'American Economic Review contient plus de texte qu'une page dans le Journal of Political Economy ou le Quarterly Journal of Economics. D'autre part, les journaux ont des longueurs moyennes d'articles différentes. Cependant, il est crucial de remarquer que les pondérations de la qualité des journaux utilisées s'entendent pour un article moyen du journal. Ainsi, ces pondérations sont censées déjà prendre en compte ces différences de typographie et de longueur moyenne des articles et il n'est donc pas nécessaire de corriger à nouveau cette caractéristique. Typiquement, le score relatif d'Economics Letters et de l'American Economic Review prend en compte le fait que le premier journal publie des articles plus courts. En revanche, parmi les articles d'un même journal, nous supposons qu'un article est plus long uniquement s'il recèle plus d'idées et d'innovations. Pour cette raison, les approches Clm et Clh corrigent la mesure de la qualité d'un article par son nombre de pages ramené au nombre de pages moyen, pour l'année de publication, des articles du journal. Ainsi, un article de nombre de pages égal au nombre moyen de pages des articles du journal de la même année a une qualité égale à la qualité du journal. S'il est 20% plus long, sa qualité est considérée comme 20% supérieure. Notons que la correction effectuée revient à modifier de moins de 50% l'indice de qualité du journal pour 80% des articles et que les français, en ayant tendance à publier des articles légèrement plus longs que la moyenne, sont plutôt avantagés par cette pondération.

Finalement, il existe un très grand nombre de façons de pondérer la qualité des journaux scientifiques. Afin de disposer d'un outil à jour et couvrant l'ensemble des journaux d'Econlit, Combes et Linnemer (2010) font un tour d'horizon des stratégies possibles et construisent plusieurs indices de qualité des journaux. Ceux-ci reposent uniquement sur des moyennes de citations reçues par les journaux, lorsque les indices de citations correspondant existent, soit environ 300 journaux (un quart d'Econlit). Lorsqu'ils n'existent pas, un modèle économétrique prédit quel serait l'indice de citation du journal. Nous retenons in fine deux indices de la qualité des journaux, prenant plus (Clh) ou moins (Clm) en compte les disparités de qualité, telle qu'appréhendée par les citations qu'ils reçoivent, des journaux. De manière importante, la hiérarchie des journaux est la même pour Clm et Clh.

La différence entre les deux vient uniquement du degré d'inégalité (convexité) des poids. Les deux systèmes de pondération n'excluent aucun journal recensé par Econlit, tous les journaux ont un poids strictement positif, alors que la plupart des classements de journaux en excluent une grande majorité (typiquement, les indices de citations considèrent au maximum un quart des journaux d'Econlit). Finalement, notons que la qualité du journal est évaluée en 2008 même pour les articles publiés une autre année.

Notons que les chercheurs français publient toujours beaucoup dans des journaux francophones. La *Revue Économique* est la plus utilisée avec 192 articles pour les chercheurs de notre base (elle compte pour 5.7% du stock 2004-2008 des articles des chercheurs actifs en 2008). Les chercheurs de notre base ont publiés quasiment 55% des 354 articles de la *Revue Économique* sur cette période. Les huit journaux les plus fréquemment utilisés sont tous francophones. Les 15 journaux les plus utilisés cumulent un tiers des publications. Parmi eux, seuls *Economics Bulletin* et *Economics Letters* ne sont pas francophones. Parmi les journaux suivants en anglais, on trouve le *Journal of Economic Theory*, le *Journal of Public Economics*, l'*European Economic Review*, le *Journal of Mathematical Economics* et *Economic Theory*. Ces journaux reflètent une certaine préférence des chercheurs français pour la formalisation. Toutefois, ces cinq journaux représentent ensemble légèrement moins d'articles que ceux de la *Revue Économique*.

En ce qui concerne la prise en compte du temps, au moins deux approches sont possibles pour décrire les publications d'un centre de recherche. La première, dite de flux, attribue à une institution les publications de l'année t de ses membres de l'année t. Cette mesure est éventuellement passée si t ne correspond pas à la date de réalisation de l'étude, ou cumulée, si l'on somme sur plusieurs années. À la date de l'étude, il est possible que certains chercheurs de l'institution n'en soient plus membres, un phénomène d'autant plus important que l'on recule dans le temps ou que l'on somme sur de nombreuses années. Pour un décideur souhaitant allouer un budget à un centre donné ou pour un étudiant souhaitant choisir un lieu où effectuer sa thèse, cette mesure n'est pas forcément la plus pertinente.

La deuxième approche, retenue ici, ne s'intéresse pas tant à la production issue de l'université mais au stock de capital recherche présent à une date donnée. Le stock à la date t consiste en la somme des publications passées (éventuellement en limitant la période de temps) des membres présents dans l'institution à la date t, indépendamment de la localisation de ces chercheurs au moment de la publication. Notons que cette approche permet aussi de suivre l'évolution des institutions au court du temps, à partir des effectifs de ces institutions à différentes dates. Pouvoir s'appuyer sur de telles listes, et non sur les affiliations référencées dans Econlit, est également un avantage pour deux raisons. D'une part, la qualité des affiliations déclarées dans Econlit est très mauvaise (en tout cas pour la France), d'autre part, elle ne permet pas d'identifier les chercheurs ne publiant pas, rendant impossible le calcul des scores par chercheur de l'institution. Plus la période est courte, plus les approches flux et stocks se rejoignent, elles seraient mêmes identiques pour la dernière année en l'absence des délais de publications.

Pour revenir au choix de la période de temps proprement dit, de nombreux classements utilisent les publications des cinq dernières années afin de mettre l'accent sur la recherche récente. Ce choix est d'ailleurs prudent lorsque des flux de publications sont mesurés. Dans notre approche en termes de stock, nous notons T=5 ans, ce choix. Nous mesurons aussi un stock fondé sur les publications de toutes les années incluses dans d'Econlit, de 1969 à

2008, ce que nous notons T=All.

Pour une longue période de recensement des publications la question du poids donné à chaque année se pose. Par analogie avec un stock de capital, une publication se déprécie dans le temps. Des publications récentes sont davantage susceptibles d'indiquer que l'institution est à la frontière de la recherche que des publications remontant au début des années soixante-dix. Cette idée est capturée de manière assez brutale lorsque seules les cinq dernières années sont considérées. Une manière plus naturelle de le faire est d'introduire un facteur de dépréciation. Cela donne notre troisième façon d'appréhender l'âge des publications que nous notons T=Dégressif. Nous utilisons un facteur d'escompte qui obéit à une fonction logistique $T\left(t(e)\right) = \frac{1-\exp\left(-10/(2009-t(e))^{1.8}\right)}{1+\exp\left(-20/(2009-t(e))^{1.2}\right)}$ dont la valeur pour les 20 premières années est donné dans le Tableau 1.1. Ainsi, un article de 2008 compte pour 1, un article de 2007 pour 0.943 articles, un de 2006 pour 0.746 etc.

TABLE 1.1 – Décompte dans le temps pour T=Dégressif

Année	Coef.	Année	Coef.
2008	1.000	1998	0.094
2007	0.943	1997	0.079
2006	0.746	1996	0.067
2005	0.549	1995	0.058
2004	0.402	1994	0.050
2003	0.299	1993	0.044
2002	0.227	1992	0.039
2001	0.177	1991	0.035
2000	0.141	1990	0.031
1999	0.114	1989	0.028

Coef. = coefficient de décompte en fonction du temps pour T=Dégressif.

Enfin, nous considérons une dernière période de temps très différente des autres, prenant en compte l'âge des chercheurs. Elle consiste à diviser le score d'un chercheur par un nombre d'années de carrière, nous la notons T=Carrière. Dans nos travaux antérieurs, faute d'information relative à l'âge des chercheurs, nous utilisions le nombre d'années depuis la première publication. Cela pouvait poser des problèmes d'interprétation pour les chercheurs publiant sur le tard pour lesquels la période de recherche était considérée comme courte et donc ceux-ci très productifs en moyenne. L'idéal est de considérer l'âge académique, à savoir le temps écoulé depuis le doctorat. Nous ne disposons pas de cette variable et tous les chercheurs ne sont en fait pas titulaires d'un doctorat. Nous utilisons donc ici l'âge des chercheurs moins 26 ans (l'âge minimum dans la base étant 27 ans). Si au niveau individuel ces différentes façons de procéder pour appréhender la durée de carrière du chercheur peuvent créer de grandes différences, nous n'en avons pas constaté au niveau agrégé des centres/universités. Notons finalement que ramener la mesure de publication à ce nombre d'années peut s'effectuer aussi bien pour T=5 ans, T=All ou T=Dégressif. Néanmoins, nous ne l'appliquons ici qu'à T=All afin de limiter le nombre de variantes.

Les mesures de publications d'un chercheur donné étant définies, nous pouvons discuter la façon dont celles-ci vont être agrégées par institution. Comme dit plus haut, notre approche nous permet de définir $\alpha_i(j)$ la part des publications du chercheur i qui va être affectée à l'institution (centre ou université) j. Le stock de publications $Y_j\left(T,W\right)$ de l'institution j pour un système de pondération W des journaux et un système T de pondération des années est donnée par

$$Y_j\left(T,W\right)=\sum_i \alpha_i(j)y_i\left(T,W\right) \tag{1.2}$$

où $y_i\left(T,W\right)$ est le stock de publications du chercheur i explicité dans l'équation (1) ci-dessus.

A ce stock est associé un nombre équivalent de chercheurs temps plein dans l'institution,

$$N_j = \sum_i \alpha_i(j) \tag{1.3}$$

Cela conduit à une mesure de stock de publications par chercheur ($Y_j^{pc}(T, W)$) de l'institution j

$$Y_j^{pc}(T, W) = \frac{Y_j(T, W)}{N(j)}. \tag{1.4}$$

1.2.3. Les indices de citations Google Scholar

L'étape préliminaire nécessaire à notre travail consiste en l'obtention des citations telles que recensées par GS reçues par chaque publication des 2832 chercheurs de notre base. Cela a été effectué via la mise au point d'une routine informatique permettant cette extraction en quelques heures. Ce point est important puisque le nombre de citations GS évolue de jour en jour sans qu'aucun historique ne soit accessible. Nous utilisons ici une extraction effectuée en janvier 2010, soit environ 2 ans après la date à laquelle nous souhaitons mesurer l'impact des centres et universités, ce qui est raisonnable étant donné le temps nécessaire à l'émergence des citations des travaux. [3]

Deux points sont cruciaux. Tout d'abord, l'extraction se fait sur la base du nom de famille et du premier prénom, alors que lorsque nous travaillons sur la base de données Econlit, l'extraction ne porte que sur le nom de famille et l'initiale du prénom sauf dans le cas d'homonymes pour lesquels nous utilisons alors le prénom complet. GS étant beaucoup

3. Nous ignorons les entrées GS de 2009, d'une part parce que les données pour 2009 semblent être très incomplètes (à la date où l'extraction a été effectuée, janvier 2010) puisque référençant moitié moins d'entrées que 2008, d'autre part, par souci de comparabilité avec la période T=5 ans utilisée pour les approches Econlit qui s'arrête en décembre 2008.

plus large qu'Econlit, le risque de présence d'homonymes serait largement accru avec une telle stratégie, sans qu'il soit possible d'y faire quoi que ce soit, comme par exemple utiliser les affiliations (déclarées dans Econlit) pour affiner la répartition en cas d'homonymie. A l'inverse, des essais nous ont montré qu'extraire sur la base du prénom complet induit une certaine perte de travaux cités. Entre deux maux, nous avons donc choisi le second, en notant qu'il n'y a pas vraiment de raisons pour que la perte soit plus importante pour certaines institutions que pour d'autres, ce qui est le critère important pour notre étude. Notons que comme pour Econlit, pour les personnes ayant des noms doubles, ou pour les femmes mariées, dont nous connaissons donc deux noms de famille notés nom1 et nom2, l'extraction est effectuée sur toutes les variantes possibles, à savoir, "nom1, prénom", mais aussi "nom2, prénom", "nom1-nom2, prénom" et "nom2-nom1, prénom". Pour les prénoms doubles, seul le premier est conservé, ce qui en l'occurrence ne semble pas vraiment accroître le nombre d'homonymes. Une dernière remarque, plus technique quoique relativement triviale mais importante pour la réplicabilité de notre étude, est que l'extraction est faite en écrivant le critère de recherche entre guillemets afin de ne pas extraire les entrées des personnes de même nom et de prénom différent (ou le contraire).

Le deuxième élément crucial dans l'extraction des données sous GS est que nous restreignons les domaines (option “subject areas" de GS) aux catégories “Business, Administration, Finance, and Economics" et “Social Sciences, Arts, and Humanities". Il aurait été tentant de considérer tous les domaines, et ce pour au moins deux raisons. D'une part, un des intérêts de GS est de prendre en compte les travaux d'économistes en dehors du domaine de l'économie. Cela est partiellement fait ici en considérant ces deux domaines déjà très larges, mais des publications en mathématiques pures par exemple ne sont pas considérées. D'autre part, l'allocation d'une publication à un domaine donné est faite automatiquement par GS, toujours via des algorithmes de proximité de mots clés. Ainsi, une publication dont le titre inclut le mot "gravité" à de fortes chances d'être classée dans le domaine "Physics, Astronomy, and Planetary Science". Les publications en économie sur les flux gravitaires de marchandises s'y retrouve donc en général également affectées. En se restreignant à ces deux domaines, il y a donc également un risque, faible tout de même, de perdre des publications véritablement en économie. Malgré ces remarques, et bien que nous effectuions les extractions sur la base du nom et du prénom des chercheurs, il est cependant nécessaire de restreindre les domaines afin de limiter les problèmes d'homonymie. Quelques essais, qui n'ont pu être systématiques faute de temps et moyens, semblent montrer que cette combinaison des sujets et du critère de recherche donne les résultats les plus fiables sur les économistes. Finalement, notons que notre procédure a été appliquée directement sur GS mais qu'elle donne des résultats quasiment identiques à ceux obtenus grâce au logiciel "Publish or Perish" de Harzing (2007), en utilisant les mêmes critères et domaines de recherche naturellement.

In fine, l'extraction produit, pour chaque chercheur, une observation pour chacune de ses entrées recensées par GS. A chaque entrée est associée son nombre de citations sur un site académique présent sur Internet. Nous ne pouvons malheureusement pas recueillir de façon automatisée l'information relative à la source de chacune de ces citations, c'est-à-dire le détail de chacune des entrées citant. Les noms des co-auteurs, et donc leur nombre, est aussi disponible. Le détail de ceux-ci est limité à 8 auteurs, sachant qu'il a longtemps été limité à 3 pour Econlit. Pour l'économie, cette limite n'est a priori pas très pénalisante puisque 97% des publications (dans Econlit en tout cas) sont effectuées à trois auteurs ou moins. Finalement, la date de publication est aussi disponible, avec cependant un nombre relativement important de valeurs manquantes, et ce d'autant plus que les dates antérieures à 1969 sont considérées comme aberrantes et transformées en valeur manquantes. 1969 peut paraître un peu récent, ce qui correspond tout de même à quarante années de publications, mais cela permet de disposer de la même date de référence que pour Econlit. Il s'avère aussi qu'un nombre relativement conséquent d'entrées est doté d'un nombre nul de citations et qu'il est en fait alors souvent difficile pour ces entrées de déterminer s'il s'agit réellement de ce qu'un chercheur qualifierait de publication. Cela peut être par exemple un résumé dans un programme de colloque, qui, en économie du moins, est rarement cité en l'état. Il s'agit aussi souvent d'erreurs de typographie sur le titre d'autres entrées référencées par ailleurs. Plus loin, certains de nos indices ignorent donc ces entrées sans citations, qui sont également souvent celles sans date.

Le Tableau 1.2 présente le nombre d'entrées recueillies, total (soit 69460) ou en équivalent écrites seul (42448) selon la méthodologie décrite plus haut c'est à dire en divisant chacune par son nombre d'auteurs. Le tableau souligne aussi le fait que presque exactement la moitié de ces entrées ne reçoit aucune citation. Pour rappel, dans la base de données Econlit, figurent, pour le même ensemble de chercheurs français, 10154 articles équivalent écrits seul. Ainsi, même en ignorant les entrées sans citations et en restreignant les sujets couverts, GS élargit très considérablement l'ensemble des publications considérées en le multipliant par deux voire quatre selon le point de vue (parmi les 10154 articles Econlit, il est fort probable que certains n'ont pas de citations, ne comparer Econlit qu'aux entrées ayant au moins une citation est donc un minimum). Le problème de date de publication manquante est réel (presque 30% des entrées) mais pas décisif pour que nous ne puissions pas, pour certains indices en tout cas, nous en servir. Notamment, sur les entrées ayant effectivement au moins une citation seulement 10% d'entre elles n'ont pas de date référencée. Finalement, sur plus courte période (5 ans), l'écart en nombre de publications entre GS et Econlit est du même ordre de grandeur.

TABLE 1.2 – Caractéristiques des bases utilisées

		nombre	%
Tout GS		69460	100.0
	eq. seul	42448	100.0
EconLit		16113	.
	eq. seul	10154	.
Citation		34776	50.1
	eq. seul	20934	49.3
Date		49569	71.4
	eq. seul	30834	72.6
Date et cit.		30645	44.1
	eq. seul	18402	43.4
5 ans		18724	27.0
	eq. seul	10378	24.4
10 ans		33131	47.7
	eq. seul	19076	44.9
5 ans et cit.		10421	15.0
	eq. seul	5540	13.1
10 ans et cit.		20234	29.1
	eq. seul	11324	26.7
EL. 5 ans		6192	.
	eq. seul	3467	.

Nombre d'entrées et pourcentage par rapport au nombre total d'entrées. Citation = entrées avec au moins une citation ; Date = entrées avec date renseignée ; Date et cit. = entrées avec date renseignée et au moins une citation ; 5 ans et 10 ans = entrées avec date renseignée, dans les 5 et 10 dernières années respectivement ; 5 ans et cit. et 10 ans et cit. = entrées avec au moins une citation et date renseignée, dans les 5 et 10 dernières années respectivement.

Comme pour les mesures de publications via Econlit, se pose la question de la prise en considération du temps et d'autres dimensions de la publication, comme son nombre de co-auteurs. Il n'est cependant plus nécessaire de prendre en compte la qualité du support et la longueur des publications, puisque l'hypothèse qui est faite alors est que le nombre de citations reçues capte directement la qualité de chaque entrée.

En ce qui concerne le temps, il est tout d'abord important de souligner que GS ne permet pas de déterminer de façon automatisée quand une citation d'une publication a été effectuée. Ainsi, détenons nous toutes les citations des entrées telles que GS les collecte sur Internet en janvier 2010, sans que l'on sache si ce sont des citations effectuées par une publication ancienne ou récente au delà du fait qu'elle est en ligne en janvier 2010. Cette propriété de GS constitue une limite dans le sens où la littérature utilisant les citations préfère en général accorder plus de poids aux citations récentes afin de mesurer l'importance pour la recherche actuelle de la publication plus que sa valeur passée.

Cela étant dit, il reste possible de jouer sur la date des entrées elles-mêmes, comme pour les mesures Econlit, même si l'on préférerait faire les deux simultanément. Nous considérons donc comme précédemment tout d'abord une fenêtre courte toujours notée T=5 ans. Nous mesurons alors les citations GS telles que référencées en janvier 2010 de toutes les entrées dont la date se situe entre 2004 et 2008. Si e est une entrée GS de date de publication $t(e)$, on peut alors définir comme précédemment une fonction de pondération des années, $T(.)$, telle que $T\left(t(e)\right) = 1$ lorsque $t(e) \geq 2004$ et $T\left(t(e)\right) = 0$ pour les années précédentes ainsi que pour 2009. Le Tableau 1.2 montre que cette période considère environ un tiers des entrées GS (ligne T=5 ans). L'influence de plus long terme fondée sur toutes les entrées GS est aussi considérée, soit T=All. Dans ce cas, $T\left(t(e)\right) = 1$ pour toutes les dates de publications, y compris quand celles-ci sont manquantes (en excluant toujours 2009). Finalement, comme pour les mesures Econlit, nous pouvons aussi décompter progressivement les publications dans le temps, ce que nous notons T=Dégressif, à l'aide de la même fonction logistique présentée dans le Tableau 1.1. Notons qu'il s'agit moins de dire que des citations anciennes ont moins de valeur que de donner moins de poids à la production ancienne. Pour donner moins de poids aux citations anciennes, il faudrait disposer de la date à laquelle les citations sont effectuées, ce que nous n'avons pas. Ce que nous faisons revient à donner moins de poids à une entrée qui a 10 ans et a reçu 30 citations qu'à une autre qui aurait reçu également 30 citations mais qui n'aurait que 5 ans. On pourrait souhaiter, en plus, donner un poids différents en fonction de la date de citation. Si les 30 citations du premier article ont été effectuées plus récemment, bien que l'article soit plus ancien, cela lui redonne de la valeur. La contrainte de données nous oblige à donner le même poids à toutes les citations.

Notons que la littérature produit aussi des indices de citations reçues par année, ce que nous notons T=Annuel. Cela revient en fait à appliquer un taux de décompte dans le temps relativement extrême car proportionnel aux nombre d'années depuis la publication, $T\left(t(e)\right) = \frac{1}{2009-t(e)}$. Il vaut mieux lire ce chiffre comme un score moyen par an sur une période plus ou moins longue, même si nous ne l'appliquons ici qu'à l'ensemble des entrées GS à date non manquante et comprise entre 1969 et 2008 (et non à celles sur les 5 dernières années par exemple).

Finalement, nous produisons comme pour les approches Econlit des scores "par année de carrière" avec alors $T\left(t(e)\right) = \frac{1}{\text{âge chercheur moins 26 ans}}$, indépendante de $t(e)$ donc.

Nous recueillons donc pour chaque chercheur le nombre de citations reçues par chacune de ses entrées GS. Comme pour le décompte simple du nombre de publications, se pose la question de savoir comment cette production est ventilée entre ses différents auteurs. Comme pour les scores Econlit, nous effectuons une ventilation entre auteurs proportionnelle à leur nombre. Si une entrée à trois auteurs recueille 21 citations, chaque auteur se voit attribué un tiers de publication. Des exemples de calculs pour les indices plus sophistiqués de citations, comme les indices H ou G, sont donnés plus bas car ils nécessitent de préciser cette ventilation.

Nous présentons dans les Tableaux 1.5 et 1.6 les corrélations entre classements de centres effectuant ou pas cette correction. On trouvera dans Bosquet et Combes (2011) ces corrélations pour les universités et pour les chercheurs. Les différences sont significatives mais pas majeures. En l'absence d'arguments en faveur de la non-prise en compte totale du nombre d'auteurs, nous préférons donc effectuer cette correction tout au long de cette étude afin également de ne pas multiplier les variantes, déjà nombreuses. A fin de comparaison avec Econlit, le Tableau 1.3 donne le pourcentage d'entrées GS à un, deux, trois ou quatre auteurs et plus. Les parts de publications à 1 ou 3 auteurs sont remarquablement similaires entre GS et Econlit. Celles à deux auteurs sont relativement moins nombreuses dans GS, alors que celle à plus de trois auteurs le sont plus. Ce dernier résultat s'explique certainement par le fait qu'Econlit n'a longtemps pas recensé le nombre d'auteurs au delà de trois. Dans tous les cas, la proximité de ces distributions est relativement rassurante puisque par ailleurs le nombre d'entrées est bien plus important pour GS : cela ne semble pas engendrer de distorsion très notable sur le co-autorat, sans toutefois que l'on puisse déterminer si les petites différences apparaissant proviennent de la nature des travaux différente entre les deux sources ou d'une pure erreur de mesure liée à la façon dont GS extrait les auteurs des publications.

TABLE 1.3 – Nombre d'auteurs par entrée

Nb. d'auteurs	1	2	3	4 et +	Total
% GS T=All	41.5	33.6	15.0	9.9	100.0
% GS T=5 ans	32.4	36.2	19.8	11.6	100.0
% EconLit T=All	42.4	39.8	15.3	2.5	100.0
% EconLit T=5 ans	31.6	43.1	20.0	5.3	100.0

Le premier indice que nous calculons, noté $Ct_i(T)$ est le nombre total de citations reçues par les travaux du chercheur i, en équivalent reçu seul, travaux produits sur la période de temps T et référencés par GS en janvier 2010. $E(i)$ est l'ensemble des entrées GS du chercheur i, e une entrée particulière dans cet ensemble, $c(e)$ le nombre de citations qu'elle a reçues, $n(e)$ son nombre de co-auteurs et $t(e)$ sa date de publication. $Ct_i(T)$ est donné par

$$Ct_i\,(T) = \sum_{e \in E(i)} T\,(t(e))\,\frac{c(e)}{n(e)}.$$

Un même nombre de citations peut cependant être obtenu soit via un petit nombre de publications très fortement citées, soit par un plus grand nombre de publications moins citées. La profession attribue en général plus de mérite à la première situation, avec sans doute plus ou moins consciemment en tête des exemples de chercheurs ayant une influence considérable via un très petit nombre d'articles publiés. La situation extrême de chercheurs n'ayant eu qu'une publication de grand impact n'est cependant pas toujours considérée comme un idéal, un bon équilibre entre volume total de citations et qualité minimale de chaque contribution étant souvent perçu comme souhaitable. Quoiqu'il en soit, ramener le nombre de citations au nombre d'entrées donne un point de vue de l'influence des travaux d'un chercheur complémentaire de celui du nombre total de citations reçues par ses entrées. Un problème lié à l'utilisation de GS est que, comme dit plus haut, le nombre d'entrées pour un chercheur donné est une variable relativement bruitée du fait de la présence d'un grand nombre d'entrées, en général non citées, difficilement assimilables à ce que l'on appellerait une publication. Afin de calculer la sensibilité de nos mesures à cette imprécision, nous calculons deux indices de nombre de citations reçues par entrée. Le premier, $Ce_i(T)$, est le nombre de citations reçues sur la période T par entrée du chercheur, quelque soit le nombre de citations reçues par ces entrées, soit $Ct_i(T)$ divisé par le nombre total d'entrées donc. Le deuxième, $Cp_i(T)$, est le nombre de citations reçues par entrée ayant reçue au moins une citation. Autrement dit, une entrée n'ayant pas reçue de citation n'est alors pas considérée comme une publication, ce qui n'affecte pas le numérateur mais modifie le dénominateur.

Dans les deux cas, le nombre qui apparaît au dénominateur est bien le nombre d'entrées équivalent écrit seul, puisque la ventilation des citations entre auteurs ne consiste pas à supposer que chaque auteur a écrit un article lui procurant un $1/n(e)$ des citations, mais bien un $1/n(e)$ d'article ayant des citations, procurant donc à cet auteur $1/n(e)$ des citations. Notons aussi que la fonction de décompte dans le temps est également prise en compte. Si l'on se restreint aux cinq dernières années, il ne faut diviser que par le nombre d'articles publiés sur ces cinq ans et l'on fait naturellement de même avec le taux de décompte continu correspondant à T=Dégressif. En revanche, pour T=Annuel, la littérature préfère diviser par le nombre total d'articles pour produire un nombre de citations par an et par article pour l'ensemble de la production, ce que nous ferons donc aussi (même si restreindre la période aurait aussi du sens). In fine,

$$Ce_i\left(T\right) = \frac{Ct_i(T)}{\sum_{e \in E(i)} \frac{T(t(e))}{n(e)}},$$

$$Cp_i\left(T\right) = \frac{Ct_i(T)}{\sum_{e \in E(i), c(e)>0} \frac{T(t(e))}{n(e)}},$$

sauf pour T=Annuel, $T\left(t(e)\right)$ étant alors remplacé par 1 au dénominateur.

Du fait, d'une part, de la relative sensibilité des mesures de citations par entrée à l'erreur de mesure sur le nombre d'entrées et du débat sur le fait de compter ou pas celles n'ayant pas de citations, d'autre part, de la nécessité de calculer deux indices, un en volume total et un par publication, pour évaluer l'influence d'un chercheur ou d'une institution, des travaux ont essayé de proposer des indices plus synthétiques combinant ces différentes dimensions. C'est par exemple le cas du désormais fameux indice H, proposé par Hirsh (2005). Notre indice $H_i(T)$ pour le chercheur i prend la valeur h, qui est unique pour ce chercheur, si h de ses publications de la période T ont reçu au moins h citations chacune. L'indice $H_i(T)$ est croissant, au sens large, à la fois avec le nombre de publications et le nombre de citations reçues par celles-ci, comme une mesure de volume total telle que $Ct_i(T)$. Cependant, en ignorant les publications les moins citées, il donne une idée du degré moyen de qualité des publications les meilleures tout en éliminant une partie du bruit provenant de la façon dont on définit une publication.

Un défaut de l'indice H a cependant été rapidement souligné. Deux chercheurs ou institutions peuvent avoir le même indice alors que l'un d'entre eux a quelques publications nettement plus citées que l'autre. Autrement dit, l'indice H ignore la distribution interne des citations reçues par les articles contribuant effectivement à son calcul. C'est principalement afin de palier ce problème que l'indice G a été proposé par Egghe (2006). Le chercheur i a un indice $G_i(T)$ qui prend la valeur g, qui est également unique, si ses g articles de la période T les plus cités ont reçu, ensemble, g^2 citations, soit g citations en moyenne cette fois-ci. On peut montrer que pour un chercheur donné $G_i(T)$ est nécessairement supérieur à $H_i(T)$, la différence entre les deux étant liée au nombre de citations reçues en moyenne par les articles les plus cités. Pour deux chercheurs ayant le même H, celui dont la moyenne des citations des articles les plus cités est la plus importante a un indice G plus élevé. Alors que ces deux indices ignorent le bas de la distribution des citations, l'indice G présente l'avantage de mieux prendre en compte le haut de celle-ci.

A partir des définitions des indices H et G de base, il faut ensuite prendre en compte le nombre de co-auteurs. Il suffit pour cela, comme l'a proposé Schreiber (2008) et dans l'esprit de ce qui est fait pour les autres indices, de simplement bien attribuer la totalité des citations à l'article mais simplement une fraction de l'article à chaque auteur. Désormais, l'indice H n'est plus nécessairement entier mais il garde la même signification. Par exemple, un indice H de 7.5 signifie que l'auteur a publié au moins 7.5 articles équivalent écrit seuls ayant au moins 7.5 citations chacun, c'est-à-dire au moins 8.

Si la prise en compte du temps affecte la valeur des citations, comme par exemple pour T=Dégressif, l'indice H peut également être étendu. Il suffit pour cela de décompter le nombre de citations avant d'ordonner les articles par ordre décroissant de celles-ci. Finalement, on peut selon les mêmes principes calculer des indices G prenant en compte le nombre de co-auteurs et le décompte dans le temps des citations. Le Tableau 1.4 donne un exemple de ces calculs à partir d'un ensemble fictif de 11 entrées d'un chercheur ayant ou pas reçues des citations GS, dont nous calculons certains de nos indices pour les trois périodes de temps T=All, T=5 ans et T=Dégressif. La première entrée a reçu 10 citations (colonne ''cit."), n'a qu'un auteur (colonne ''aut.") et a été publiée en 2006 (colonne "date"), et de même pour les suivantes. En fonction de la date de publication, on obtient le poids lié à la période de publication, donné dans la colonne "pp". Celui-ci vaut toujours 1 pour T=All, 1 pour T=5 ans lorsque l'entrée a moins de 5 ans, 0 sinon, et les poids correspondant à chaque année donnés dans le Tableau 1.1 pour T=Dégressif. La colonne ''pc" donne le poids lié au nombre de co-auteurs. La colonne suivante "cp" agrège ces deux informations pour obtenir le nombre pondéré de citations GS qui contribue au score GS obtenu. La colonne "rp" donne finalement le rang, également pondéré, de l'entrée GS, qui est nécessaire pour calculer les indices H et G. Les cinq dernières lignes du tableau donnent les cinq scores GS principaux correspondant à cet exemple, pour chaque de ces trois périodes, que l'on peut retrouver à l'aide des formules données précédemment.

TABLE 1.4 – Exemples de calculs de scores GS

entrée	cit.	aut.	date	T=All				T=5 ans				T=Dégressif			
				pp	pc	cp	rp	pp	pc	cp	rp	pp	pc	cp	rp
1	10	1	2006	1	1	10	1	1	1	10	1	0,75	1	7,46	0,75
2	5	1	2008	1	1	5	2	1	1	5	2	1,00	1	5,00	1,75
3	5	1	2003	1	1	5	3	0	1	0	2	0,30	1	1,49	5,81
4	5	2	2007	1	0,5	5	3,5	1	0,5	5	2,5	0,94	0,5	4,72	2,22
5	3	1	2008	1	1	3	4,5	1	1	3	3,5	1,00	1	3,00	3,22
6	3	1	2006	1	1	3	5,5	1	1	3	4,5	0,75	1	2,24	3,96
7	3	1	2005	1	1	3	6,5	1	1	3	5,5	0,55	1	1,65	5,51
8	2	1	2003	1	1	2	7,5	0	1	0	5,5	0,30	1	0,60	6,11
9	2	1	2008	1	1	2	8,5	1	1	2	6,5	1,00	1	2,00	4,96
10	1	1	2003	1	1	1	9,5	0	1	0	6,5	0,30	1	0,30	6,41
11	0	1	2008	1	1	0	10,5	1	1	0	7,5	1,00	1	0,00	7,41
Ct				39.0				31.0				28.5			
Ce				3.5				3.9				3.6			
Cp				3.9				4.4				4.1			
H				3.5				2.5				2.2			
G				5.5				4.5				4.0			

Entrée= numéro entrée GS ; cit.= nombre de citations reçues ; date= date de l'entrée, aut.= nombre de co-auteurs ; pp= poids période ; pc= poids co-auteurs, cp= citations pondérées ; re= rang de l'entrée.

En résumé, nous calculons, sur cinq périodes de temps (toutes les années, T=All), les cinq dernières années (2004-2008, T=5 ans), en décomptant dans le temps (T=Dégressif), par année (d'existence de la publication, T=Annuel), et par année de carrière (du chercheur, T=Carrière), cinq indicateurs de citation GS au niveau individuel : le nombre total de citations ($Ct(T)$), le nombre de citations par entrée GS ($Ce(T)$), le nombre de citations par entrée GS ayant reçu au moins une citation ($Cp(T)$), l'indice H ($H(T)$), et l'indice G ($G(T)$), toutes prenant en compte le nombre de co-auteurs de façon proportionnelle. Naturellement, d'autres variantes sont possibles et ont d'ailleurs fait l'objet de calcul de notre part, comme le fait de ne pas prendre en compte le nombre de co-auteurs. Néanmoins, les résultats obtenus ne sont alors pas fondamentalement différents alors que la présentation en est considérablement alourdie. Ainsi, avons-nous choisi de nous concentrer sur ce groupe d'indices GS, et même, dans les sections de classements, suite aux résultats obtenus quant à la très forte corrélation des classements observées entre certains d'entre eux, sur un sous ensemble de ceux-ci permettant de mettre en valeur différentes dimensions de l'influence des publications des chercheurs.

Maintenant que nous avons défini les mesures de citations d'un chercheur donné, nous pouvons discuter la façon dont celles-ci sont agrégées par institution. Nous le faisons de la même façon que pour les approches Econlit : toutes les citations reçues par un chercheur sont affectées au centre de recherche ou à l'université dont il est membre en 2008 et pour les chercheurs à affiliations multiples, une fraction de chaque article est attribué à chaque institution. Ainsi, lorsqu'un chercheur équitablement réparti entre deux centres a une entrée GS ayant reçu 20 citations, le centre est considéré comme ayant obtenu une demi entrée GS ayant reçu 20 citations. La première étape consiste ainsi à établir la liste des entrées GS de chaque centre ou université avec son nombre de co-auteurs et la part qui lui en revient, prenant à la fois en compte la part de co-auteurs effectivement membre du centre et le possible décompte dans le temps. La stratégie de calcul d'indice appliquée aux chercheurs peut alors être appliquée exactement de la même façon aux centres ou universités.

Finalement, on peut vouloir comparer la productivité moyenne des chercheurs entre universités. Pour les indices, $Ct(T)$, $Ce(T)$, $Cp(T)$, il suffit de les diviser par le nombre de chercheurs équivalent temps plein de l'institution. En revanche, pour les indices H et G, il faut bien en faire la moyenne entre chercheurs, pondérée par la part du chercheur dans le centre.

Nous avons déjà mentionné un certain nombre de limites dues à l'imperfection de l'outil GS, qui il est vrai, reste un outil relativement jeune qui devrait s'améliorer avec les années. Certaines entrées ne devraient pas apparaître, d'autres devraient être regroupées, la date de publication de certaines est manquante, le nombre d'auteurs n'est pas toujours facile à établir. Plus important est le fait de devoir réduire quelque peu les champs couverts afin de ne pas trop multiplier les problèmes d'homonymes. Ne pas connaître la date à laquelle les citations sont reçues empêche d'aborder une autre famille d'indices prenant en compte la date à laquelle les citations ont eu lieu. Finalement, de façon plus large, connaître entièrement l'origine de la citation permettrait d'affiner les indices produits en ignorant les citations émanant des auteurs eux-mêmes par exemple, ou en pondérant plus fortement les citations provenant d'auteurs, ou de journaux, eux-même fortement cités. La structure actuelle du site internet GS ne permet pas pour le moment d'effectuer de telles variantes.

1.2.4. Corrélations entre classements

Une des conclusions de nos travaux précédents (Combes et Linnemer, 2001 ; Combes et Linnemer, 2003b ; Bosquet et al., 2010) est la relative insensibilité de la hiérarchie des centres ou universités aux hypothèses effectuées quant à la façon de prendre en compte le nombre de co-auteurs, la longueur des publications, voire la qualité des journaux. Elle est confirmée par la présente étude pour l'approche GS comme le montre les Tableaux 1.5 et 1.6.

Prendre ou pas le nombre de co-auteurs en compte n'induit quasiment pas de changement dans la hiérarchie des institutions, la corrélation la plus faible entre deux indices identiques prenant ou pas en compte ce nombre étant par exemple de 0.96 pour l'indice H au niveau des centres. Pour cette raison, nous préférons utiliser les indices qui prennent en compte ce nombre, puisque, comme expliqué plus haut, l'intuition, en termes d'hypothèses sous-jacentes et d'agrégation, est plus claire.

Pour le nombre de citations reçues par entrée, le fait d'exclure ou pas celles n'ayant aucune citation est lui aussi tout à fait insensible, avec des corrélations supérieures à 0.97 au niveau des classements des centres, entre indices qui le font ou pas, bien que la part d'entrées n'ayant pas de citation soit relativement élevée : le phénomène semble affecter tous les centres dans des proportions relativement similaires. Ces conclusions sur le type d'entrées considérées et la prise en compte du nombre de co-auteurs au niveau des centres se retrouvent non seulement au niveau des universités mais également des chercheurs, comme cela apparaît dans Bosquet et Combes (2011).

A façons de prendre en compte le nombre d'auteurs et le nombre d'entrées identiques, les corrélations sont élevées entre tous les indices fondés sur les citations GS. Il est intéressant de noter que les corrélations des indices H et G sont plus élevées avec les indices de citations totales qu'avec les indices de citations par entrée, ces dernières étant même plus faibles qu'entre les indices en volume total et par entrée. Ainsi, si l'intérêt des indices H et G est de prendre en compte la distribution des citations entre publications, ils restent néanmoins fortement corrélés au volume total de citations reçues et beaucoup moins au nombre de citations par entrée. Cela n'est pas tellement plus marqué pour l'indice H que pour l'indice G, bien que ce dernier corresponde à un nombre moyen de citations par entrée (mais également à un nombre d'entrées).

La corrélation élevée des classements en nombre de citations totales et par entrée, supérieure à celle de ces derniers avec les indices H et G, montre que les institutions recevant beaucoup de citations sont celles où la qualité de chacune des publications est elle même élevée. Recevoir beaucoup de citations ne consiste pas seulement à effectuer beaucoup de travaux mais à également avoir une qualité élevée pour chacun d'entre eux.

Une originalité de notre travail est de permettre de comparer les indices de citations (GS ici) avec des indices très différents dans leur nature, fondés sur la qualité moyenne des supports de publication et non la qualité de chaque publication prise isolément. Il s'avère que la hiérarchie des centres évaluée selon les approches GS, quelles qu'elles soient, est relativement proche de la hiérarchie obtenue via Clm, avec des corrélations de rang supérieures à 0.7 que ce soit pour les citations totales reçues ou les indices H ou G. Des corrélations légèrement plus faibles sont observées avec les scores par entrée, de façon relativement naturelle puisque les scores Clm correspondent à des volumes de publications. Ainsi, utiliser la qualité des journaux comme prédicteur de la qualité des publications, en tous les cas du nombre de citations qu'elles reçoivent, semble constituer une stratégie tout à fait pertinente. Les corrélations sont légèrement plus faibles avec Clh, quoique restant largement positives. Ainsi, les indicateurs de citations GS semblent moins discriminants que ne le sont les indicateurs Clh, et se rapprochent plus des indicateurs Clm. Notons que des explications de ce résultat peuvent se trouver dans le fait que les citations GS ne prennent pas en compte la qualité du support citant, alors que les indicateurs de citations des journaux à partir desquels Clm et Clh sont construits le font. Ces corrélations sont légèrement moins élevées en ce qui concerne la hiérarchie des centres selon leurs scores par chercheur, ce qui s'explique sans doute par la plus grande variabilité globale de ceux-ci.

Les résultats observés au niveau des centres sont confirmés au niveau des universités, les classements desquelles sont en général encore plus corrélés que ceux des centres. Notam-

TABLE 1.5 – Corrélations entre rangs des centres, scores totaux, T=Dégressif

	Ct1	Ct	Ce1	Ce	Cp1	Cp	H1	H	G1	G	Clm	Clh
Ct1	1	0.99	0.86	0.87	0.85	0.85	0.96	0.95	0.97	0.98	0.84	0.73
Ct		1	0.86	0.87	0.85	0.86	0.95	0.96	0.95	0.97	0.85	0.72
Ce1			1	0.99	0.98	0.98	0.79	0.79	0.82	0.83	0.75	0.63
Ce				1	0.97	0.98	0.81	0.82	0.83	0.84	0.78	0.65
Cp1					1	1	0.78	0.78	0.80	0.81	0.73	0.59
Cp						1	0.79	0.79	0.80	0.82	0.74	0.61
H1							1	0.96	0.93	0.95	0.85	0.75
H								1	0.92	0.93	0.86	0.73
G1									1	0.99	0.82	0.75
G										1	0.83	0.75
Clm											1	0.91
Clh												1

Ct = citations totales ; Ce = citations par entrée ; Cp = citations par entrée ayant au moins une citation ; H = Indice H ; G = Indice G ; 1 indique que le nombre de co-auteurs n'est pas pris en compte. Clm et Clh correspondent aux indices Econlit plus (Clh) ou moins (Clm) sélectifs.

TABLE 1.6 – Corrélations entre rangs des centres, scores par chercheur, T=Dégressif

	Ct1	Ct	Ce1	Ce	Cp1	Cp	H1	H	G1	G	Clm	Clh
Ct1	1	0.98	0.82	0.81	0.82	0.82	0.94	0.92	0.95	0.93	0.72	0.66
Ct		1	0.80	0.80	0.79	0.79	0.93	0.94	0.95	0.94	0.72	0.65
Ce1			1	0.98	0.94	0.94	0.80	0.77	0.77	0.73	0.63	0.60
Ce				1	0.92	0.94	0.80	0.78	0.77	0.73	0.63	0.61
Cp1					1	0.99	0.78	0.75	0.78	0.74	0.57	0.55
Cp						1	0.79	0.76	0.78	0.74	0.58	0.56
H1							1	0.97	0.97	0.96	0.75	0.67
H								1	0.95	0.97	0.74	0.64
G1									1	0.97	0.73	0.66
G										1	0.71	0.61
Clm											1	0.92
Clh												1

Ct = citations totales ; Ce = citations par entrée ; Cp = citations par entrée ayant au moins une citation ; H = Indice H ; G = Indice G ; 1 indique que le nombre de co-auteurs n'est pas pris en compte. Clm et Clh correspondent aux indices Econlit plus (Clh) ou moins (Clm) sélectifs.

ment, les corrélations des scores en volume de citations ou volume Clm sont supérieures à 0.79, légèrement inférieures pour les volumes par chercheur. En désagrégeant l'information au contraire, c'est-à-dire en se plaçant cette fois-ci au niveau des chercheurs, les corrélations entre classements GS restent très élevées, seulement très légèrement plus faibles qu'entre centres et universités. Les corrélations entre scores GS et scores Clm, et encore plus Clh, deviennent plus faibles, ce qui reflète un résultat relativement standard relatif au fait que l'agrégation par centre ou université tend à rendre la hiérarchie obtenue relativement moins sensible à la méthode utilisée. Les corrélations entre scores en volume restent néanmoins toutes supérieures à 0.6. Le détail de ces corrélations aux niveaux universités et chercheurs peut être trouvé dans Bosquet et Combes (2011).

Finalement, les Tableaux 1.7 et 1.8 présentent également des corrélations entre centres, mais cette fois-ci lorsque la période d'étude est modifiée. Là encore, l'on est surtout frappé par les valeurs très élevées prises par ces corrélations. L'ordre des centres est donc remarquablement stable selon les différentes périodes d'étude, et il en va de même pour les universités et les chercheurs comme le montrent les corrélations présentées dans Bosquet et Combes (2011).

TABLE 1.7 – Corrélations des rangs des centres entre périodes, scores totaux

	Ct					G			
	T=All	T=5 ans	T=An.	T=Dég.	T=Car.	T=All	T=5 ans	T=An.	T=Dég.
Ct, T=All	1	0.92	0.98	0.97	0.97	0.96	0.87	0.89	0.92
Ct, T=5 ans		1	0.97	0.98	0.95	0.91	0.97	0.93	0.96
Ct, T=Annuel			1	1	0.98	0.96	0.93	0.93	0.96
Ct, T=Dégressif				1	0.97	0.95	0.94	0.93	0.97
Ct, T=Carrière					1	0.94	0.92	0.91	0.95
G, T=All						1	0.92	0.95	0.97
G, T=5 ans							1	0.96	0.98
G, T=Annuel								1	0.99
G, T=Dégressif									1

Ct = citations totales ; G = Indice G.

TABLE 1.8 – Corrélations des rangs des centres entre périodes, scores par chercheur

	Ct					G			
	T=All	T=5 ans	T=An.	T=Dég.	T=Car.	T=All	T=5 ans	T=An.	T=Dég.
Ct, T=All	1	0.85	0.96	0.94	0.93	0.95	0.85	0.91	0.89
Ct, T=5 ans		1	0.94	0.96	0.90	0.87	0.92	0.90	0.90
Ct, T=Annuel			1	0.99	0.95	0.94	0.92	0.96	0.94
Ct, T=Dégressif				1	0.95	0.93	0.93	0.95	0.94
Ct, T=Carrière					1	0.90	0.87	0.91	0.89
G, T=All						1	0.92	0.97	0.95
G, T=5 ans							1	0.94	0.98
G, T=Annuel								1	0.96
G, T=Dégressif									1

Ct = citations totales ; G = Indice G.

1.3. Disparités de citations et tendances nationales

1.3.1. Citations Google Scholar de la recherche en économie

Comme l'indique le Tableau 1.9, les chercheurs localisés en France reçoivent ensemble quasiment 265000 citations GS en janvier 2010, 47000 pour leurs entrées de la période 2004-2008. Chaque entrée a en moyenne presque 7 citations, le double pour celles en ayant au moins une. La France a un indice H d'environ 180, soit 180 entrées ayant au moins 180 citations, sachant que, comme nous le verrons plus loin, le chercheur au dernier centile a un indice H d'environ 17, et la France a environ 500 entrées ayant en moyenne 500 citations (indice G). Sur 5 ans, ces deux indices sont divisés par un peu plus que deux. La France obtient environ 34000 citations GS par an, et une nouvelle entrée obtient en moyenne plus d'une citation par année, 2 une fois qu'elle en a obtenue au moins une. Elle a environ 50 entrées recevant au moins 50 citations par an en moyenne.

Le Tableau 1.10 indique que sur l'ensemble de leur carrière, seuls 6.2% des chercheurs n'ont pas d'entrée GS et 15% n'en ont pas ayant au moins une citation, alors que 27.1% n'ont pas de publication Econlit. Sur les cinq dernières années, alors que 45.3% n'ont pas de

TABLE 1.9 – Scores France entière

	Ct	Ce	Cp	H	G
T=All	264897.7	6.7	13.9	177.9	486.1
T=5 ans	46568.0	4.8	9.1	72.9	217.4
T=Annuel	34246.1	1.2	2.0	48.2	134.8
T=Dégressif	50809.3	5.6	10.4	50.3	218.5

publication Econlit, seulement 22.4% n'ont pas d'entrée GS et 34.3% n'en ont pas avec au moins une citation. Ainsi, la production accrue des français en termes d'entrées GS n'est pas simplement due au fait que les chercheurs publiant selon Econlit ont plus d'entrées GS que de publications Econlit mais aussi au fait que la part des non publiant est plus faible selon GS. Cela est relativement "normal", puisque le support est élargi, mais nous disposons donc ici d'une mesure précise de l'élargissement produit. On trouvera dans Bosquet et Combes (2011) ces parts de publiant pour chacune des universités de notre étude.

TABLE 1.10 – Parts de chercheurs publiant, France entière

	Pub. GS	Cit. GS	Pub. EL
T=All	93.8	85.0	72.9
T=5 ans	77.6	65.7	54.7

Etant donné ces résultats, ainsi que les corrélations obtenues entre classements, nous nous concentrons principalement dans la suite de cet article sur deux indices GS relativement différents, à savoir le nombre de citations totales et l'indice G, qui sont systématiquement comparés à Clm.

1.3.2. Disparités entre chercheurs, centres et universités

Les disparités de publication des chercheurs sont grandes, ce qui constitue un fait général relativement ancien et solidement établi depuis Lotka (1926). Plus intéressant est donc la possibilité de comparer le degré de disparités selon l'outil de mesure utilisé, ce qui correspond à l'un des objets de cet article. Le Tableau 1.11 compare pour la période T=Dégressif un certain nombre d'indicateurs de disparités. Il s'avère que l'indice G est relativement peu discriminant, en tous les cas bien moins que le nombre total de citations reçues, ce qui apparaît pour quasiment toutes les mesures de disparités calculées ici. Par exemple, alors que le rapport inter-quartiles entre universités est de 2 pour l'indice G, il est de quasiment 13 en termes de citations totales, et l'on obtient des ratios de 5 et 32 pour les rapports inter-déciles. Les écarts sont similaires au niveau des centres.

Il est également intéressant de noter que l'indice Clm est en fait intermédiaire en termes de disparité entre les indicateurs de nombre total de citations reçues et l'indice G. Cela est vrai non seulement sur le haut de la distribution (rapport P90/P50) mais également sur le bas (rapport P50/P10). Cela est plus surprenant, puisque, au niveau des chercheurs, Clm semble peu discriminant sur le bas de la distribution, un grand nombre d'entre eux ayant en fait un score nul, alors que les scores GS continuent à décroître de façon relativement continue. On trouvera dans Bosquet et Combes (2011) des tableaux identiques pour T=All et T=5 ans à partir desquels des conclusions similaires sont obtenues. Dans la section 1.3.3 décrivant la distribution d'un grand nombre de scores, des comparaisons supplémentaires de disparités des différents indices sont données.

Un dernier indicateur de concentration de la production intra-institution est donné par la part des citations du chercheur le plus cité de l'institution (Tableau 1.12). L'approche

TABLE 1.11 – Indicateurs de dispersion des scores, T=Dégressif

	université									**centre**									**chercheur**		
	inter t.			**inter c.**			**intra**			**inter t.**			**inter c.**			**intra**					
	Ct	G	Clm	Ct	G	Clm	Ct	G	Clm	Ct	G	Clm	Ct	G	Clm	Ct	G	Clm	Ct	G	Clm
coef. var.	2.4	0.6	1.2	2.2	0.8	0.9	1.9	1.2	1.6	2.1	0.7	1.4	2.0	0.8	1.2	1.8	1.2	1.6	4.2	1.3	2.5
p75/p25	12.8	2.0	3.0	5.4	2.7	2.7	.	.	.	6.3	2.1	3.3	4.9	2.3	3.6	.	.	.	21.0	7.7	.
p90/p10	31.5	5.2	12.4	27.2	5.7	8.0	.	.	.	29.8	5.2	12.4	27.0	4.7	9.7	.	.	.	.	.	.
p50/p10	5.1	2.4	3.1	4.1	2.3	2.6	.	.	.	4.7	2.4	2.9	4.1	2.0	2.9	.	.	.	.	.	.
p90/p50	6.2	2.2	4.1	6.7	2.4	3.1	.	.	.	6.3	2.2	4.2	6.6	2.4	3.3	.	.	.	10.4	4.1	14.3

"inter t." : dispersion entre universités ou centres des scores totaux, "inter c." : dispersion entre universités ou centres des scores par chercheur, "intra" : dispersion intra-universités ou centres, coeff. var. : coefficient de variation, p75/p25 : rapport inter-quartile, p90/p10 : rapport dernier/premier décile, p50/p10 : rapport médiane/premier décile, p90/p50 : rapport dernier décile/médiane.

GS, avec des parts moyennes de ce chercheur aux alentours de 40%, conduit à un poids encore plus élevé de celui-ci dans la production de son centre que l'approche Econlit, ce que l'on retrouve en tout point de la distribution de cette grandeur.

TABLE 1.12 – Distribution de la part du chercheur le plus cité dans le centre (%)

	moy.	c.v.	P1	P5	P10	P25	P50	P75	P90	P95	P99
Ct, T=All	37.2	0.5	9.6	14.8	17.2	21.8	33.5	46.3	64.4	69.9	86.2
Ct, T=5 ans	41.0	0.5	9.7	14.2	18.0	25.4	33.2	53.0	75.0	79.2	91.7
Clm, T=All	29.5	0.6	8.3	11.1	13.4	18.4	24.6	37.4	47.9	58.3	83.5
Clm, T=5 ans	34.4	0.6	5.9	10.1	15.9	20.7	28.5	42.5	57.7	83.5	100.0

Ct = citations totales ; Clm est un indice de publication Econlit. Lecture : le chercheur qui est le plus cité dans chaque laboratoire pèse en moyenne 37,2% des citations totales du laboratoire. Il y a 10% des laboratoires français de recherche en économie où le meilleur chercheur pèse moins de 17,2% des citations.

1.3.3. Distributions des scores des chercheurs

Cette section présente la distribution des scores des chercheurs, l'intérêt étant aussi de permettre à un chercheur de se situer dans la distribution nationale.[4] Le caractère plus ou moins sélectif des différentes approches apparaît clairement. Sont présentés dans

4. Sans oublier que le nombre de citations GS est en expansion constante et que la présente étude se base sur des données de janvier 2010.

les Tableau 1.13 et 1.14 les distributions pour l'ensemble des citations reçues pour T=All et T=5 ans respectivement, celles pour T=Dégressif et T=Carrière étant données dans Bosquet et Combes (2011).

En termes de disparités, il est bien confirmé que si l'indice Clm est plus sélectif que les indices H, G ou par entrée, il l'est moins que ceux en nombre total de citations et pourrait constituer à ce titre une bonne synthèse entre ceux-ci. Cela n'est plus le cas pour Clh qui est le plus sélectif de tous. On observe que les indices GS permettent de mieux discriminer le bas de la distribution pour lequel des scores nuls, et donc non discriminants, sont obtenus selon Clm et Clh. Le nombre total de citations reçues est très sélectif. Il est divisé par un facteur 4 lorsque l'on passe du dernier centile au 95ème, puis encore d'un facteur deux en passant de celui-ci au dernier décile. Il y a un facteur 100 entre le chercheur médian et celui situé au dernier centile, pour un facteur 70 en Clm, et de l'ordre de 8 et 12 respectivement pour les indices H et G. L'ensemble de tous les travaux du chercheur médian ne représente un stock de citations équivalent reçues seul que de 14.67. Il a 2 articles ayant reçu au moins 2 citations (indice H) et 4 qui en ont reçu en moyenne 4. Le chercheur au dernier centile a un stock d'environ 17 articles ayant reçu au moins 17 citations, mais tout de même 48 articles en ayant reçu en moyenne 48.

Sur les autres périodes de temps, les scores obtenus deviennent rapidement très faibles, comme le montre le Tableau 1.14 pour T=5 ans et les tableaux donnés dans Bosquet et Combes (2011) pour T=Dégressif et T=Carrière, reflétant le fait qu'être cité nécessite en tout état de cause du temps, ne serait-ce que par les délais de publications des articles citant même si ce temps devrait être réduit pour des citations sur Internet. Le chercheur au dernier centile bénéficie tout de même de 232 citations équivalent reçues seuls en janvier

2010 pour ses travaux des 5 années 2004-2008, ses 6 articles de la période les plus cités sont cités au moins 6 fois et il a environ 16 articles cités en moyenne 16 fois.

TABLE 1.13 – Distributions des scores des chercheurs, T=All

index	mean	c.v.	P10	P15	P20	P25	P50	P75	P90	P95	P99
Ct1	170.87	5.87	0	1.00	2.00	4.00	24.00	99.00	341.00	633.00	2378.00
Ct	97.22	6.27	0	0.20	1.00	2.50	14.67	53.17	179.67	349.43	1387.92
Ce1	3.79	1.80	0	0.07	0.40	0.64	2.00	4.38	9.07	13.24	25.93
Ce	3.63	1.86	0	0.03	0.39	0.60	1.87	4.27	8.37	12.70	27.00
Cp1	6.99	1.30	0	1.00	1.50	2.00	4.67	8.80	15.47	21.82	38.34
Cp	6.73	1.33	0	1.00	1.50	2.00	4.50	8.50	14.82	20.94	39.40
H1	3.94	1.21	0	1.00	1.00	1.00	3.00	5.00	9.00	13.00	23.00
H	2.76	1.21	0	0.17	0.50	0.83	2.00	3.58	6.33	8.67	16.85
G1	7.13	1.41	0	1.00	1.00	2.00	4.00	9.00	17.00	23.00	48.00
G	6.96	1.39	0	0.20	1.00	1.33	4.00	8.83	16.42	22.83	48.20
Clm	53.84	3.53	0	0	0	0	11.40	40.38	109.70	205.84	771.55
Clh	13.73	7.46	0	0	0	0	0.05	0.76	11.82	43.12	370.31

c.v. = coefficient of variation ; Ct = citations totales ; Ce = citations par entrée ; Cp = citations par entrée ayant au moins une citation ; H = Indice H ; G = Indice G ; 1 indique que le nombre de co-auteurs n'est pas pris en compte. Clm et Clh sont les indices Econlit plus (Clh) ou moins (Clm) sélectifs.Variables are defined in section 3.2. To match what is done in the econometric section, publication variables are first computed as three-year forward moving averages before descriptive statistics are computed. The number of observations for panel (a), (b) and (c) are 38,577, 12,591, and 1209, respectively. Some individuals have been deleted from Table 3.1 because of missing values. Descriptive statistics at the department level (panel (c) are calculated on the sub-sample of departments in which there is at least one publisher and hence, for which all variables are defined. That is why departments average size is slightly higher than in Table 3.1.

TABLE 1.14 – Distributions des scores des chercheurs, T=5 ans

index	mean	c.v.	P10	P15	P20	P25	P50	P75	P90	P95	P99
Ct1	33.95	3.55	0	0	0	0	4.00	23.00	83.00	142.00	424.00
Ct	17.23	3.78	0	0	0	0	2.50	12.00	38.30	68.83	232.07
Ce1	2.78	1.82	0	0	0	0	1.00	3.36	7.38	11.20	24.25
Ce	2.75	1.91	0	0	0	0	1.00	3.22	7.11	10.91	25.29
Cp1	4.40	1.72	0	0	0	0	2.00	5.75	11.00	16.00	35.00
Cp	4.37	1.79	0	0	0	0	2.00	5.50	10.81	15.71	34.52
H1	1.75	1.23	0	0	0	0	1.00	2.00	4.00	6.00	10.00
H	1.11	1.26	0	0	0	0	0.75	1.75	2.83	3.67	6.12
G1	2.95	1.46	0	0	0	0	2.00	4.00	8.00	11.00	20.00
G	2.41	1.48	0	0	0	0	1.17	3.33	6.57	9.00	15.98
Clm	17.81	2.48	0	0	0	0	2.56	16.92	48.97	86.95	195.06
Clh	3.38	5.68	0	0	0	0	0.01	0.19	3.40	16.07	74.32

c.v. = coefficient of variation ; Ct = citations totales ; Ce = citations par entrée ; Cp = citations par entrée ayant au moins une citation ; H = Indice H ; G = Indice G ; 1 indique que le nombre de co-auteurs n'est pas pris en compte. Clm et Clh sont les indices Econlit plus (Clh) ou moins (Clm) sélectifs.

1.3.4. Citations en fonction de l'âge

La distribution des âges des chercheurs (particulière par sa bi-modalité illustrée dans Bosquet et al. (2010) avec une sur-représentation des chercheurs d'environ 38 et 60 ans) est susceptible d'évoluer sensiblement dans le temps. De plus, les différentes générations de chercheurs de notre base n'ont pas fait face aux mêmes incitations à publier et aux mêmes supports pour le faire. Nous présentons, néanmoins, ici les différences de publication des chercheurs français en fonction de leur âge. Afin de rendre la comparaison pertinente et contourner les effets de stocks, nous sommes obligés de nous placer sur une période identique pour tous et relativement courte, soit T=5 ans, même si les résultats pour les autres périodes sont donnés dans Bosquet et Combes (2011).

Le résultat marquant est que l'on retrouve une courbe en cloche de la productivité des chercheurs pour les indices GS, quels qu'ils soient, comme on l'obtient pour les indices Clm et Clh. Le pic de productivité n'est pas atteint tout à fait au même âge selon les indices utilisés, le plus tardif étant observé pour les citations totales alors que pour les citations par article il est un des plus précoces : les chercheurs jeunes semblent privilégier la qualité pour chaque article, les plus âgés ayant plus tendance à produire aussi des articles à diffusion plus restreinte. Pour les indices H et G, cette variation au cours du cycle de vie des chercheurs est moins marquée, alors qu'elle est quasiment maximale pour les indices Clm et Clh.

1.3.5. Disparités entre statuts

Les chercheurs français n'ont pas tous le même statut avec des temps consacrés à l'activité de publication d'articles de recherche différents. Les Tableaux 1.16 et 1.17 mesurent

TABLE 1.15 – Scores par chercheur par classe d'âge, T=5 ans

	26-30	31-35	36-40	41-45	46-50	51-55	56-60	>60
Ct1	19.71	32.77	36.02	40.05	44.29	50.09	19.78	24.55
Ct	10.43	16.82	17.79	18.81	21.90	25.76	10.71	13.45
Ce1	2.52	3.54	3.35	3.28	2.60	2.44	1.71	2.22
Ce	2.33	3.45	3.35	3.25	2.50	2.43	1.69	2.24
Cp1	4.11	5.39	5.08	5.37	4.12	3.71	2.75	3.87
Cp	4.02	5.30	5.02	5.30	4.04	3.68	2.74	3.94
H1	1.73	2.12	2.07	1.89	1.88	1.77	1.29	1.21
H	1.18	1.40	1.27	1.18	1.15	1.10	0.85	0.78
G1	2.75	3.45	3.38	3.28	3.27	3.11	2.07	2.16
G	2.38	2.89	2.67	2.60	2.66	2.58	1.79	1.80
Clm	16.62	24.05	22.19	23.06	20.56	16.32	9.25	9.29
Clh	1.70	2.78	4.17	5.54	5.04	4.61	1.07	1.28

Ct = citations totales ; Ce = citations par entrée ; Cp = citations par entrée ayant au moins une citation ; H = Indice H ; G = Indice G ; 1 indique que le nombre de co-auteurs n'est pas pris en compte. Clm et Clh sont les indices Econlit plus (Clh) ou moins (Clm) sélectifs.

respectivement les citations totales et indices G en volume et par chercheur pour chacun des statuts de notre base en indiquant les variations par rapport à la hiérarchie selon Clm. Les statuts apparaissent en gras italique s'ils progressent d'au moins 3 rangs, et en italique simple s'ils régressent d'au moins 3 rangs entre les deux types d'approches, la variation exacte étant donnée entre parenthèses.

Il apparaît que les quatre ou cinq statuts les plus productifs selon les indices GS restent identiques à ceux identifiés au moyen de Clm et Clh. En revanche, il est intéressant de noter que progressent selon les mesures GS les statuts qui ne sont pas au coeur de notre analyse, comme peuvent l'être les maîtres de conférence de la section 5 ou les chargés de recherche de la section 37. Cela s'explique relativement aisément par le fait que GS élargit le spectre de publications considérées, notamment en direction d'autres disciplines, connexes à l'économie tout de même comme la gestion, la sociologie, l'histoire ou la géographie etc. Ainsi, les "assimilés" professeurs ou chargés de recherche, en poste dans des grandes écoles, parfois de commerce ou dans des administrations, les chercheurs Cnrs non section

37 connaissent-ils les progressions les plus importantes.

La hiérarchie des statuts est très proche lorsqu'on la mesure à l'aide de l'indice G moyen ou du nombre de citations par chercheur (colonne de droite des Tableaux 1.16 et 1.17). Elle varie plus lorsqu'on compare le nombre de citations totales reçues et l'indice G de l'ensemble du statut (colonne de gauche des Tableaux 1.16 et 1.17), les écarts se resserrant d'une part et les statuts étant très productifs en moyenne remontant en termes de G grâce, probablement, à leurs meilleurs éléments.

Table 1.16 – Statuts, Citations totales, T=Dégressif

statut	rg.	tot.	nor.
PR 05 (567)(0)	1	16711.2	100.0
MCF 05 (1281)(0)	2	7252.0	43.4
DR cnrs (84)(0)	3	5597.6	33.5
Ponts et C. (7)(+1)	4	4254.8	25.5
A-PR (97)(+1)	5	3826.8	22.9
PR non 05 (158)(+2)	6	3406.6	20.4
DE ehess (13)(+2)	7	2654.0	15.9
CR cnrs (102)(-4)	8	2299.2	13.8
DR inra (56)(+1)	9	1493.2	8.9
Insee (22)(+2)	10	1368.9	8.2
CR inra (89)(-4)	11	1268.7	7.6
MCF non 05 (260)(-1)	12	1080.6	6.5
A-CR (21)(0)	13	743.2	4.5
DR cnrs non 37 (10)(+5)	14	253.3	1.5
Autre (13)(+2)	15	249.8	1.5
Insee non EC (27)(-2)	16	242.1	1.5
CR cnrs non 37 (9)(+1)	17	157.9	0.9
A-MCF (10)(-3)	18	88.3	0.5
Prag (6)(-3)	19	59.7	0.4

statut	rg.	p.c.	nor.
Ponts et C. (7)(0)	1	571.11	100.0
DE ehess (13)(0)	2	208.16	36.5
DR cnrs (84)(0)	3	66.60	11.7
Insee (22)(0)	4	60.98	10.7
A-PR (97)(+4)	5	39.64	6.9
A-CR (21)(+4)	6	35.64	6.2
PR 05 (567)(-1)	7	29.47	5.2
DR inra (56)(-1)	8	26.66	4.7
DR cnrs non 37 (10)(+10)	9	25.33	4.4
CR cnrs (102)(-5)	10	22.56	4.0
PR non 05 (158)(+1)	11	21.58	3.8
Autre (13)(+5)	12	19.21	3.4
CR cnrs non 37 (9)(+5)	13	17.54	3.1
CR inra (89)(-6)	14	14.26	2.5
Prag (6)(-4)	15	9.95	1.7
A-MCF (10)(-3)	16	9.19	1.6
Insee non EC (27)(-2)	17	8.97	1.6
MCF 05 (1281)(-4)	18	5.66	1.0
MCF non 05 (260)(-3)	19	4.16	0.7

La colonne "rg." donne le rang, la colonne "tot." donne le score total, "nor." le score normalisé par rapport à celui du premier classé, "p.c." donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

Le Tableau 1.18 corrige de la durée de la carrière. Les statuts apparaissant ci-dessus les plus productifs sont en effet aussi ceux pour lequel l'âge moyen est le plus élevé, ce qui a donné aux chercheurs plus de temps pour publier et donc pour recevoir des citations. Il

TABLE 1.17 – Statuts, Indice G, T=Dégressif

statut	rg.	tot.	nor.
PR 05 (567)(0)	1	112.7	100.0
A-PR (97)(+4)	2	87.9	78.0
DR cnrs (84)(0)	3	85.1	75.5
MCF 05 (1281)(-2)	4	79.6	70.6
PR non 05 (158)(+3)	5	79.5	70.5
Ponts et C. (7)(-1)	6	75.0	66.5
DE ehess (13)(+2)	7	73.4	65.1
Insee (22)(+4)	8	62.2	55.2
CR cnrs (102)(-5)	9	56.9	50.5
DR inra (56)(0)	10	51.8	46.0
CR inra (89)(-4)	11	46.7	41.5
A-CR (21)(+1)	12	42.3	37.5
MCF non 05 (260)(-2)	13	33.6	29.8
Insee non EC (27)(0)	14	21.6	19.2
Autre (13)(+2)	15	21.3	18.9
DR cnrs non 37 (10)(+3)	16	18.1	16.0
CR cnrs non 37 (9)(+1)	17	12.8	11.4
A-MCF (10)(-3)	18	12.7	11.3
Prag (6)(-3)	19	9.9	8.8

statut	rg.	p.c.	nor.
Ponts et C. (7)(0)	1	13.90	100.0
DE ehess (13)(0)	2	12.53	90.2
DR cnrs (84)(0)	3	6.71	48.3
Insee (22)(0)	4	5.19	37.4
DR inra (56)(+2)	5	4.81	34.6
DR cnrs non 37 (10)(+13)	6	4.33	31.2
CR cnrs (102)(-2)	7	4.13	29.7
Autre (13)(+9)	8	4.05	29.2
PR 05 (567)(-3)	9	3.72	26.8
A-PR (97)(-1)	10	3.51	25.3
A-CR (21)(-1)	11	3.02	21.7
PR non 05 (158)(0)	12	2.95	21.2
CR inra (89)(-5)	13	2.75	19.8
CR cnrs non 37 (9)(+4)	14	2.64	19.0
A-MCF (10)(-2)	15	2.36	17.0
Prag (6)(-5)	16	2.09	15.0
Insee non EC (27)(-2)	17	1.70	12.2
MCF 05 (1281)(-4)	18	1.41	10.1
MCF non 05 (260)(-3)	19	1.19	8.6

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

apparaît que les différences observées entre statuts ne sont cependant pas dues à ce simple effet de l'âge, puisque la hiérarchie entre statuts est alors très proche de celle décrite précédemment.

Finalement, le Tableau 1.19 hiérarchise les statuts selon leur nombre de citations reçues par entrée GS. Il s'agit d'un point de vue relativement différent de celui des tableaux précédents qui correspond vraiment à une efficacité par article diffusé. Les statuts recevant beaucoup de citations globalement en reçoivent également beaucoup par entrée, mais on voit que là encore, les statuts qui ne sont pas au cœur de notre analyse progressent par rapport à Clm parfois encore plus que précédemment.

TABLE 1.18 – Statuts, Citations totales, T=Carrière

statut	rg.	tot.	nor.
PR 05 (567)(0)	1	3830.6	100.0
MCF 05 (1281)(0)	2	2014.1	52.6
DR cnrs (84)(0)	3	1425.3	37.2
Ponts et C. (7)(+1)	4	1286.3	33.6
A-PR (97)(+1)	5	913.7	23.9
PR non 05 (158)(+3)	6	874.0	22.8
DE ehess (13)(+1)	7	654.1	17.1
CR cnrs (102)(-4)	8	634.9	16.6
CR inra (89)(-2)	9	432.6	11.3
Insee (22)(+1)	10	398.4	10.4
DR inra (56)(-1)	11	324.7	8.5
MCF non 05 (260)(0)	12	301.7	7.9
A-CR (21)(0)	13	170.4	4.5
Autre (13)(+1)	14	59.6	1.6
DR cnrs non 37 (10)(+4)	15	53.1	1.4
CR cnrs non 37 (9)(0)	16	42.7	1.1
Insee non EC (27)(-1)	17	21.1	0.6
A-MCF (10)(-4)	18	16.3	0.4
Prag (6)(-1)	19	7.0	0.2

statut	rg.	p.c.	nor.
Ponts et C. (7)(0)	1	172.66	100.0
DE ehess (13)(0)	2	51.30	29.7
Insee (22)(+1)	3	17.74	10.3
DR cnrs (84)(-1)	4	16.96	9.8
A-PR (97)(+3)	5	9.46	5.5
A-CR (21)(+4)	6	8.17	4.7
PR 05 (567)(-1)	7	6.76	3.9
CR cnrs (102)(-3)	8	6.23	3.6
DR inra (56)(0)	9	5.80	3.4
PR non 05 (158)(+2)	10	5.54	3.2
DR cnrs non 37 (10)(+8)	11	5.31	3.1
CR inra (89)(-5)	12	4.86	2.8
CR cnrs non 37 (9)(+2)	13	4.74	2.8
Autre (13)(+2)	14	4.58	2.7
A-MCF (10)(-4)	15	1.70	1.0
MCF 05 (1281)(-3)	16	1.57	0.9
Prag (6)(-3)	17	1.17	0.7
MCF non 05 (260)(-1)	18	1.16	0.7
Insee non EC (27)(-1)	19	0.78	0.5

La colonne "rg." donne le rang, la colonne "tot." donne le score total, "nor." le score normalisé par rapport à celui du premier classé, "p.c." donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.4. Principaux classements

Nous présentons dans cette section nos principaux classements. La période de temps retenue est T=Dégressif, ce qui représente un bon compromis avec la perspective de long terme, peut-être encore plus cruciale pour les indices de citations, celles-ci nécessitant nécessairement plus de temps pour se révéler, et une perspective de plus court-terme sûrement plus importante lorsque l'on se projette dans l'avenir.

Dans tous les tableaux de cette section (et de manière plus générale pour presque tous les tableaux de classements) la structure est la suivante. Le sous-tableau de gauche donne le classement en volume total, celui de droite le classement par chercheur. Dans chaque

TABLE 1.19 – Statuts, Citations par entrée, T=Dégressif

statut	rg.	tot.	nor.
MCF 05 (1281)(+1)	1	3105.9	100.0
PR 05 (567)(-1)	2	2590.6	83.4
A-PR (97)(+3)	3	836.3	26.9
PR non 05 (158)(+4)	4	757.4	24.4
DR cnrs (84)(-2)	5	573.5	18.5
MCF non 05 (260)(+5)	6	573.2	18.5
CR cnrs (102)(-3)	7	377.6	12.2
CR inra (89)(-1)	8	315.8	10.2
DR inra (56)(+1)	9	241.0	7.8
Insee (22)(+2)	10	166.8	5.4
Ponts et C. (7)(-6)	11	159.0	5.1
A-CR (21)(+1)	12	149.0	4.8
DE ehess (13)(-4)	13	144.1	4.6
DR cnrs non 37 (10)(+5)	14	143.8	4.6
Insee non EC (27)(-1)	15	87.6	2.8
Autre (13)(+1)	16	43.2	1.4
A-MCF (10)(-2)	17	32.5	1.0
CR cnrs non 37 (9)(0)	18	30.2	1.0
Prag (6)(-3)	19	14.1	0.5

statut	rg.	p.c.	nor.
Ponts et C. (7)(0)	1	21.34	100.0
DR cnrs non 37 (10)(+17)	2	14.38	67.4
DE ehess (13)(-1)	3	11.30	53.0
A-PR (97)(+5)	4	8.66	40.6
Insee (22)(-1)	5	7.43	34.8
A-CR (21)(+4)	6	7.15	33.5
DR cnrs (84)(-4)	7	6.82	32.0
PR non 05 (158)(+4)	8	4.80	22.5
PR 05 (567)(-3)	9	4.57	21.4
DR inra (56)(-3)	10	4.30	20.2
CR cnrs (102)(-6)	11	3.71	17.4
CR inra (89)(-4)	12	3.55	16.6
A-MCF (10)(0)	13	3.38	15.8
CR cnrs non 37 (9)(+4)	14	3.36	15.8
Autre (13)(+2)	15	3.32	15.6
Insee non EC (27)(-1)	16	3.24	15.2
MCF 05 (1281)(-3)	17	2.42	11.4
Prag (6)(-7)	18	2.35	11.0
MCF non 05 (260)(-3)	19	2.20	10.3

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

sous-tableau, les premières parenthèses après le nom de l'université/centre indiquent son nombre de chercheurs équivalent temps plein. Entre les deuxièmes parenthèses se trouve la variation du classement par rapport au classement Clm. Si cette variation est supérieure ou égale à 3, le nom de l'université/centre est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras. La colonne “rg.” donne le rang de l'université/centre, la colonne “tot.” (resp. “p.c.”) son score (resp. son score par chercheur), enfin, la colonne “nor.” indique le score normalisé en % du score de la meilleure université.

La section 1.4.1 rassemble quatre classements des universités. D'abord, un classement en volume et un par chercheur pour les citations totales, puis ces deux mêmes classements pour l'indice G. La section 1.4.2 présente les mêmes classements au niveau des centres de recherche.

Ces quatre classements apportent des éclairages différents. Tout d'abord, les classements en termes de volume mettent en avant les universités/centres dont la visibilité globale est la plus grande. Les classements en termes de scores par chercheur peuvent toutefois présenter des différences et ils permettent de mettre en avant des universités/centres avec un petit nombre de chercheurs très productifs mais dont la petite taille limite le rang dans les classements en volume. Certaines universités/centres sont bien classés selon les deux approches.

Nous ne re-détaillons pas les résultats obtenus dans toutes les sections de classements, chaque lecteur pouvant facilement chercher les éléments qui l'intéressent. Néanmoins, deux conclusions importantes semblent se dégager. D'une part, pour une majorité de centres ou universités, la variation de classements par rapport à Clm est faible. D'autre part, pour quelques autres, elles est relativement forte, et il s'avère que ces unités semblent être celles dont le cœur de l'activité n'est pas l'économie, ou, ce qui est lié, dont les membres ne sont pas au cœur de notre champ, à savoir les sections 5 du Cnu et 37 du Cnrs.

1.4.1. Classements des universités

1.4.1.1. Citations totales

TABLE 1.20 – Universités, Citations totales, T=Dégressif

université	rg.	tot.	nor.
Pse-Paris 1 (214)(+1)	1	9731.6	100.0
Tse-Toulouse 1 (125)(-1)	2	9550.6	98.1
Crest-Ensae (67)(+1)	3	3487.8	35.8
Hec (75)(+1)	4	2874.1	29.5
Aix Marseille 2-3 (114)(-2)	5	2233.8	23.0
Paris 9 (124)(+6)	6	1553.0	16.0
Nancy 2-Strasbourg 1 (95)(-1)	7	1262.5	13.0
Ec. Polytechnique (33)(0)	8	1216.0	12.5
Paris 10 (80)(-2)	9	1212.9	12.5
Grenoble 2-Inra (128)(+4)	10	980.2	10.1
Lille 1-Polytech Lille (153)(-1)	11	963.9	9.9
Montpellier 1-Inra (62)(+1)	12	914.5	9.4
Lyon 2 (70)(+3)	13	877.8	9.0
Iep Paris (9)(+5)	14	860.3	8.8
Cergy Pontoise (37)(-6)	15	789.0	8.1
Dijon (65)(+14)	16	750.2	7.7
Bordeaux 4 (72)(-2)	17	717.4	7.4
Nice (83)(-1)	18	697.7	7.2
Caen-Rennes 1 (121)(-8)	19	695.1	7.1
Clermont 1 (32)(+3)	20	536.4	5.5
Paris 11 (36)(+21)	21	525.9	5.4
Inra Vers-Grig (12)(+2)	22	517.4	5.3
Paris 13 (45)(+2)	23	514.1	5.3
Paris 2 (41)(-6)	24	457.2	4.7
Cired (14)(+7)	25	451.1	4.6
Besancon (24)(-4)	26	402.3	4.1
Inra Ivry (37)(-1)	27	347.2	3.6
Inra Rennes (12)(-1)	28	343.1	3.5
Versailles St Quentin (24)(+12)	29	319.6	3.3
Nantes (23)(-9)	30	288.3	3.0
Strasbourg 3 (13)(-2)	31	281.9	2.9
Ens Cachan (7)(+1)	32	251.7	2.6
Orleans (34)(-5)	33	234.5	2.4
Lille 2 (13)(+12)	34	233.5	2.4
Chambery (15)(+16)	35	206.9	2.1
Le Mans (18)(-16)	36	204.4	2.1
Pau (17)(+7)	37	202.3	2.1
Reims (32)(-3)	38	197.5	2.0
Paris 8 (27)(-5)	39	185.5	1.9
Evry (18)(-9)	40	169.3	1.7
St Etienne (22)(-3)	41	145.4	1.5
Paris 12 (30)(+13)	42	139.6	1.4
Inra Dijon (11)(-7)	43	125.9	1.3
Paris 7 (10)(+26)	44	114.7	1.2
Rouen (22)(+19)	45	106.9	1.1
Limoges (18)(-1)	46	96.5	1.0
Lille 3 (11)(-4)	47	88.4	0.9
La Reunion (19)(-9)	48	87.4	0.9
Perpignan (12)(-12)	49	85.3	0.9
Inra Nancy (7)(+15)	50	84.0	0.9
Angers (18)(+8)	51	73.3	0.8
Montpellier 3 (10)(+2)	52	70.6	0.7
La Rochelle (5)(0)	53	69.9	0.7
Toulon (11)(+3)	54	63.2	0.7
Poitiers (27)(-15)	55	61.3	0.6
Tours (12)(-9)	56	59.2	0.6

université	rg.	p.c.	nor.
Iep Paris (9)(+1)	1	101.21	100.0
Tse-Toulouse 1 (125)(-1)	2	76.65	75.7
Crest-Ensae (67)(0)	3	51.98	51.4
Pse-Paris 1 (214)(+1)	4	45.52	45.0
Inra Vers-Grig (12)(+2)	5	43.11	42.6
Hec (75)(+5)	6	38.37	37.9
Ec. Polytechnique (33)(-3)	7	36.41	36.0
Ens Cachan (7)(-2)	8	34.33	33.9
Cired (14)(+6)	9	33.41	33.0
Inra Rennes (12)(-1)	10	28.59	28.3
Strasbourg 3 (13)(+3)	11	22.02	21.8
Cergy Pontoise (37)(-4)	12	21.32	21.1
Aix Marseille 2-3 (114)(-1)	13	19.53	19.3
Lille 2 (13)(+17)	14	17.96	17.8
Besancon (24)(+3)	15	16.76	16.6
Clermont 1 (32)(+10)	16	16.76	16.6
Paris 10 (80)(+3)	17	15.15	15.0
Montpellier 1-Inra (62)(+8)	18	14.75	14.6
Paris 11 (36)(+47)	19	14.61	14.4
La Rochelle (5)(-7)	20	13.98	13.8
Chambery (15)(+26)	21	13.79	13.6
Nancy 2-Strasbourg 1 (95)(0)	22	13.36	13.2
Versailles St Quentin (24)(+32)	23	13.32	13.2
Inra Nancy (7)(+17)	24	12.92	12.8
Nantes (23)(-8)	25	12.81	12.7
Paris 9 (124)(+22)	26	12.48	12.3
Lyon 2 (70)(+5)	27	12.47	12.3
Pau (17)(+14)	28	12.26	12.1
Dijon (65)(+36)	29	11.56	11.4
Paris 13 (45)(+10)	30	11.55	11.4
Paris 7 (10)(+37)	31	11.47	11.3
Inra Dijon (11)(-16)	32	11.45	11.3
Paris 2 (41)(-9)	33	11.29	11.2
Le Mans (18)(-24)	34	11.20	11.1
Bordeaux 4 (72)(-1)	35	10.03	9.9
Evry (18)(-15)	36	9.41	9.3
Inra Ivry (37)(-1)	37	9.38	9.3
Nice (83)(+14)	38	8.46	8.4
Lille 3 (11)(-16)	39	8.03	7.9
Rennes 2 (7)(+11)	40	7.73	7.6
Grenoble 2-Inra (128)(+19)	41	7.66	7.6
Montpellier 3 (10)(-14)	42	7.43	7.4
Perpignan (12)(-24)	43	7.41	7.3
Orleans (34)(-7)	44	6.90	6.8
Paris 8 (27)(-7)	45	6.87	6.8
St Etienne (22)(-2)	46	6.61	6.5
Valenciennes (8)(+14)	47	6.40	6.3
Lille 1-Polytech Lille (153)(+6)	48	6.29	6.2
Reims (32)(+1)	49	6.27	6.2
Mulhouse (9)(-23)	50	6.16	6.1
Cnam (7)(-16)	51	6.05	6.0
Caen-Rennes 1 (121)(-9)	52	5.77	5.7
Toulon (11)(-7)	53	5.75	5.7
Toulouse 2 (10)(+21)	54	5.49	5.4
Limoges (18)(-6)	55	5.36	5.3
Marne La Vallee (11)(-27)	56	5.08	5.0

suite page suivante

suite de la page précédente

université	**rg.**	**tot.**	**nor.**
Littoral (13)(-9)	57	54.2	0.6
Rennes 2 (7)(+9)	58	54.1	0.6
Marne La Vallee (11)(-7)	59	53.3	0.6
Mulhouse (9)(-4)	60	52.4	0.5
Toulouse 2 (10)(+14)	61	52.2	0.5
Brest (21)(-13)	62	51.2	0.5
Antilles Guyane (22)(-13)	63	50.5	0.5
Valenciennes (8)(+5)	64	48.0	0.5
Le Havre (15)(+1)	65	42.5	0.4
Cnam (7)(-5)	66	42.4	0.4
Artois (13)(-7)	67	39.8	0.4
Metz (13)(-5)	68	38.6	0.4
Paris 5 (11)(-11)	69	33.1	0.3
Amiens (22)(-8)	70	29.7	0.3
Toulouse 3 (10)(+2)	71	27.7	0.3
Lyon 1 (5)(-1)	72	24.5	0.3
Corte (9)(-1)	73	15.9	0.2
Bretagne Sud (9)(-6)	74	13.3	0.1
Lyon 3 (9)(-1)	75	8.1	0.1

université	**rg.**	**p.c.**	**nor.**
Tours (12)(-27)	57	4.93	4.9
Lyon 1 (5)(-5)	58	4.91	4.9
Rouen (22)(+12)	59	4.86	4.8
Paris 12 (30)(+9)	60	4.73	4.7
La Reunion (19)(-22)	61	4.60	4.6
Littoral (13)(-29)	62	4.17	4.1
Angers (18)(+1)	63	4.07	4.0
Artois (13)(-6)	64	3.19	3.2
Paris 5 (11)(-20)	65	3.16	3.1
Metz (13)(-4)	66	3.09	3.1
Le Havre (15)(0)	67	2.83	2.8
Toulouse 3 (10)(+5)	68	2.77	2.7
Brest (21)(-13)	69	2.50	2.5
Antilles Guyane (22)(-11)	70	2.35	2.3
Poitiers (27)(-14)	71	2.27	2.2
Corte (9)(0)	72	1.77	1.8
Bretagne Sud (9)(-10)	73	1.48	1.5
Amiens (22)(-4)	74	1.35	1.3
Lyon 3 (9)(-1)	75	0.90	0.9

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.4.1.2. Indice G

TABLE 1.21 – Universités, Indice G, T=Dégressif

université	**rg.**	**tot.**	**nor.**
Pse-Paris 1 (214)(+1)	1	112.0	100.0
Tse-Toulouse 1 (125)(-1)	2	104.6	93.4
Hec (75)(+2)	3	72.7	64.9
Crest-Ensae (67)(0)	4	64.2	57.4
Aix Marseille 2-3 (114)(-2)	5	63.2	56.5
Nancy 2-Strasbourg 1 (95)(0)	6	52.4	46.8
Paris 9 (124)(+5)	7	47.2	42.2
Ec. Polytechnique (33)(0)	8	45.6	40.8
Paris 10 (80)(-2)	9	45.1	40.3
Iep Paris (9)(+9)	10	43.6	38.9
Montpellier 1-Inra (62)(+2)	11	43.1	38.5
Lyon 2 (70)(+4)	12	40.4	36.1
Grenoble 2-Inra (128)(+1)	13	36.4	32.5
Lille 1-Polytech Lille (153)(-3)	14	36.4	32.5
Paris 2 (41)(+3)	15	36.0	32.2
Nice (83)(+1)	16	35.0	31.2
Inra Vers-Grig (12)(+7)	17	34.0	30.4
Dijon (65)(+12)	18	32.4	29.0
Caen-Rennes 1 (121)(-8)	19	32.1	28.7
Cergy Pontoise (37)(-10)	20	32.1	28.7
Le Mans (18)(-1)	21	31.0	27.7
Cired (14)(+10)	22	30.7	27.5
Clermont 1 (32)(0)	23	30.0	26.8
Bordeaux 4 (72)(-9)	24	29.7	26.5
Paris 11 (36)(+17)	25	29.4	26.3
Strasbourg 3 (13)(+3)	26	27.2	24.3
Inra Rennes (12)(0)	27	26.2	23.4
Ens Cachan (7)(+5)	28	25.8	23.0
Paris 13 (45)(-4)	29	24.8	22.1

université	**rg.**	**p.c.**	**nor.**
Iep Paris (9)(+1)	1	9.23	100.0
Cired (14)(+13)	2	5.53	59.9
Inra Vers-Grig (12)(+4)	3	5.34	57.9
Tse-Toulouse 1 (125)(-3)	4	5.07	55.0
Inra Rennes (12)(+4)	5	4.95	53.6
Ec. Polytechnique (33)(-2)	6	4.83	52.3
Pse-Paris 1 (214)(-2)	7	4.58	49.6
Ens Cachan (7)(-2)	8	4.35	47.1
Crest-Ensae (67)(-6)	9	4.31	46.7
Hec (75)(+1)	10	3.51	38.0
Inra Nancy (7)(+30)	11	3.46	37.5
Cergy Pontoise (37)(-4)	12	3.36	36.4
Strasbourg 3 (13)(+1)	13	3.24	35.1
Chambery (15)(+33)	14	3.22	34.9
Clermont 1 (32)(+10)	15	2.96	32.1
Aix Marseille 2-3 (114)(-4)	16	2.84	30.7
Paris 10 (80)(+3)	17	2.75	29.8
Paris 13 (45)(+22)	18	2.74	29.7
Paris 11 (36)(+47)	19	2.65	28.7
La Rochelle (5)(-7)	20	2.48	26.9
Perpignan (12)(-2)	21	2.43	26.3
Besancon (24)(-4)	22	2.39	25.8
Lyon 2 (70)(+9)	23	2.36	25.5
Versailles St Quentin (24)(+32)	24	2.36	25.6
Nancy 2-Strasbourg 1 (95)(-3)	25	2.35	25.5
Le Mans (18)(-16)	26	2.33	25.3
Montpellier 3 (10)(+1)	27	2.32	25.2
Nantes (23)(-11)	28	2.30	24.9
Rennes 2 (7)(+22)	29	2.27	24.6

suite page suivante

suite de la page précédente

université	**rg.**	**tot.**	**nor.**
Besancon (24)(-8)	30	24.5	21.9
Inra Ivry (37)(-5)	31	24.4	21.8
Versailles St Quentin (24)(+9)	32	24.0	21.5
Lille 2 (13)(+13)	33	23.7	21.2
Chambery (15)(+17)	34	22.8	20.3
Nantes (23)(-14)	35	22.2	19.9
Pau (17)(+8)	36	20.9	18.6
Reims (32)(-2)	37	20.3	18.1
Evry (18)(-7)	38	18.7	16.7
Orleans (34)(-11)	39	18.2	16.2
Limoges (18)(+5)	40	17.2	15.4
Inra Dijon (11)(-5)	41	17.1	15.3
Paris 7 (10)(+28)	42	16.3	14.6
St Etienne (22)(-5)	43	15.9	14.2
Paris 12 (30)(+11)	44	15.3	13.7
Angers (18)(+14)	45	15.2	13.6
Lille 3 (11)(-3)	46	14.4	12.9
Inra Nancy (7)(+18)	47	13.8	12.3
Paris 8 (27)(-14)	48	13.5	12.0
Rouen (22)(+15)	49	13.2	11.8
La Reunion (19)(-11)	50	13.1	11.7
Valenciennes (8)(+18)	51	13.0	11.7
Montpellier 3 (10)(+2)	52	12.9	11.5
Perpignan (12)(-16)	53	12.5	11.1
Toulon (11)(+3)	54	12.2	10.9
Mulhouse (9)(+1)	55	11.9	10.7
Poitiers (27)(-16)	56	11.8	10.5
Tours (12)(-10)	57	11.7	10.5
Artois (13)(+2)	58	11.1	9.9
Brest (21)(-9)	59	11.1	9.9
Marne La Vallee (11)(-8)	60	10.4	9.3
Toulouse 2 (10)(+15)	61	10.4	9.3
Le Havre (15)(+4)	62	10.1	9.0
Metz (13)(0)	63	9.7	8.7
Antilles Guyane (22)(-14)	64	9.6	8.6
Cnam (7)(-4)	65	9.4	8.4
La Rochelle (5)(-13)	66	9.3	8.3
Rennes 2 (7)(+1)	67	9.3	8.3
Paris 5 (11)(-10)	68	8.1	7.2
Littoral (13)(-21)	69	7.6	6.8
Toulouse 3 (10)(+3)	70	5.9	5.2
Corte (9)(+1)	71	5.2	4.7
Lyon 3 (9)(+3)	72	5.2	4.7
Lyon 1 (5)(-2)	73	4.6	4.1
Amiens (22)(-12)	74	4.5	4.0
Bretagne Sud (9)(-7)	75	4.2	3.8

université	**rg.**	**p.c.**	**nor.**
Evry (18)(-9)	30	2.21	23.9
Inra Dijon (11)(-15)	31	2.20	23.8
Montpellier 1-Inra (62)(-6)	32	2.17	23.5
Pau (17)(+10)	33	2.17	23.5
Dijon (65)(+31)	34	2.15	23.3
Lille 3 (11)(-12)	35	2.12	23.0
Bordeaux 4 (72)(-2)	36	2.03	22.0
Inra Ivry (37)(-1)	37	2.01	21.7
Paris 9 (124)(+11)	38	2.01	21.8
Cnam (7)(-4)	39	1.92	20.8
Grenoble 2-Inra (128)(+20)	40	1.84	19.9
Nice (83)(+11)	41	1.82	19.7
Paris 2 (41)(-18)	42	1.78	19.3
Paris 7 (10)(+25)	43	1.74	18.8
Caen-Rennes 1 (121)(-1)	44	1.73	18.7
St Etienne (22)(0)	45	1.73	18.7
Reims (32)(+4)	46	1.68	18.2
Lille 2 (13)(-16)	47	1.67	18.1
Orleans (34)(-11)	48	1.64	17.8
Paris 8 (27)(-11)	49	1.62	17.5
Toulon (11)(-3)	50	1.62	17.5
Marne La Vallee (11)(-22)	51	1.56	16.9
Lille 1-Polytech Lille (153)(+2)	52	1.51	16.3
Mulhouse (9)(-26)	53	1.49	16.2
Toulouse 2 (10)(+22)	54	1.49	16.2
Rouen (22)(+16)	55	1.42	15.4
La Reunion (19)(-17)	56	1.36	14.8
Littoral (13)(-24)	57	1.34	14.5
Paris 12 (30)(+11)	58	1.29	14.0
Le Havre (15)(+8)	59	1.28	13.8
Tours (12)(-30)	60	1.21	13.1
Limoges (18)(-12)	61	1.19	12.8
Toulouse 3 (10)(+11)	62	1.01	10.9
Valenciennes (8)(-1)	63	1.01	11.0
Paris 5 (11)(-19)	64	0.93	10.1
Lyon 1 (5)(-12)	65	0.92	9.9
Metz (13)(-4)	66	0.88	9.5
Poitiers (27)(-9)	67	0.88	9.5
Brest (21)(-12)	68	0.83	9.0
Corte (9)(+4)	69	0.83	9.0
Antilles Guyane (22)(-11)	70	0.81	8.8
Angers (18)(-7)	71	0.78	8.5
Bretagne Sud (9)(-9)	72	0.77	8.3
Artois (13)(-15)	73	0.75	8.1
Amiens (22)(-4)	74	0.63	6.8
Lyon 3 (9)(-1)	75	0.45	4.9

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.4.2. Classements des centres

TABLE 1.22 – Centres en 2008, Citations totales, T=Dégressif

centre	rg.	tot.	nor.
Gremaq (Tse-Toulouse 1) (59)(0)	1	7575.5	100.0
Pjse (Pse-Paris 1) (43)(+1)	2	5530.9	73.0
Ces (Pse-Paris 1) (138)(-1)	3	4009.1	52.9
Greghec (Hec) (75)(+1)	4	2874.1	37.9
Greqam (Aix Marseille 2-3) (45)(-1)	5	1903.9	25.1
Non-Grecsta *(Crest-Ensae)* (36)(+6)	6	1777.0	23.5
Grecsta (Crest-Ensae) (31)(-1)	7	1710.8	22.6
Lerna (Tse-Toulouse 1) (19)(-1)	8	1296.1	17.1
Preg (Ec. Polytechnique) (33)(0)	9	1216.0	16.1
Economix (Paris 10) (66)(-2)	10	1186.5	15.7
Beta (Nancy 2-Strasb. 1) (71)(-1)	11	1185.4	15.7
Drm *(Paris 9)* (69)(+15)	12	1093.7	14.4
Centre *(Iep Paris)* (9)(+5)	13	860.3	11.4
Thema (Cergy Pontoise) (36)(-3)	14	787.8	10.4
Lameta (Montpellier 1-Inra) (33)(+1)	15	737.2	9.7
Gredeg (Nice) (71)(+1)	16	634.3	8.4
Crem (Caen-Rennes 1) (82)(-4)	17	586.8	7.8
Cerdi *(Clermont 1)* (26)(+7)	18	536.4	7.1
Gate (Lyon 2) (27)(-4)	19	534.9	7.1
Eco. Pub. *(Inra Vers-Grig)* (12)(+6)	20	517.4	6.8
Leg *(Dijon)* (42)(+26)	21	478.8	6.3
Non-Drm (Paris 9) (56)(-8)	22	459.3	6.1
Centre *(Cired)* (14)(+14)	23	451.1	6.0
Cepn *(Paris 13)* (33)(+7)	24	438.7	5.8
Lem (Lille 1-Poly. Lille) (87)(-2)	25	435.5	5.8
Gretha (Bordeaux 4) (37)(-6)	26	434.4	5.7
Centre (Besancon) (24)(-3)	27	402.3	5.3
Autre (Tse-Toulouse 1) (23)(-7)	28	375.2	5.0
Cerag *(Grenoble 2-Inra)* (48)(+25)	29	355.5	4.7
Clerse-Eco *(Lille 1-Poly. Lille)* (36)(+4)	30	352.4	4.7
Smart (Inra Rennes) (12)(-1)	31	343.1	4.5
Non-Gate *(Lyon 2)* (43)(+14)	32	342.9	4.5
Centre *(Vers. St Quentin)* (24)(+20)	33	319.6	4.2
Ermes (Paris 2) (17)(-1)	34	298.3	3.9
Centre (Nantes) (23)(-13)	35	288.3	3.8
Non-Gretha *(Bordeaux 4)* (35)(+5)	36	283.0	3.7
Centre (Strasbourg 3) (13)(-2)	37	281.9	3.7
Gael (Grenoble 2-Inra) (15)(+2)	38	273.5	3.6
Non-Cermes *(Paris 11)* (16)(+24)	39	272.6	3.6
Cermes *(Paris 11)* (21)(+59)	40	253.3	3.3
Lirhe-Eco *(Tse-Toulouse 1)* (20)(+44)	41	252.5	3.3
Centre (Ens Cachan) (7)(-4)	42	251.7	3.3
Centre *(Lille 2)* (13)(+17)	43	233.5	3.1
Leo (Orleans) (30)(-12)	44	231.6	3.1
Lepii *(Grenoble 2-Inra)* (29)(+29)	45	231.0	3.1
Iredu-Eco *(Dijon)* (12)(+46)	46	212.5	2.8
Centre *(Chambery)* (15)(+19)	47	206.9	2.7
Centre (Le Mans) (18)(-29)	48	204.4	2.7
Centre *(Pau)* (17)(+7)	49	202.3	2.7
Autre *(Aix Marseille 2-3)* (57)(+7)	50	202.2	2.7
Centre (Reims) (32)(-7)	51	197.5	2.6
Autre *(Pse-Paris 1)* (33)(+16)	52	191.6	2.5
Centre (Paris 8) (27)(-14)	53	185.5	2.5
Aliss (Inra Ivry) (19)(-25)	54	177.0	2.3
Autre (Lille 1-Poly. Lille) (31)(-27)	55	176.1	2.3
Mona-Tsv *(Inra Ivry)* (19)(+45)	56	170.2	2.3
Centre (Evry) (18)(-21)	57	169.3	2.2
Non-Ermes (Paris 2) (24)(-16)	58	158.9	2.1
Centre (St Etienne) (22)(-9)	59	145.4	1.9
Centre *(Paris 12)* (30)(+11)	60	139.6	1.8

centre	rg.	p.c.	nor.
Pjse (Pse-Paris 1) (43)(+1)	1	128.18	100.0
Gremaq (Tse-Toulouse 1) (59)(-1)	2	127.96	99.8
Centre (Iep Paris) (9)(+2)	3	101.21	79.0
Lerna (Tse-Toulouse 1) (19)(-1)	4	70.06	54.7
Grecsta (Crest-Ensae) (31)(+1)	5	54.83	42.8
Non-Grecsta *(Crest-Ensae)* (36)(+6)	6	49.50	38.6
Eco. Pub. (Inra Vers-Grig) (12)(+2)	7	43.11	33.6
Greqam (Aix Marseille 2-3) (45)(-4)	8	42.45	33.1
Greghec *(Hec)* (75)(+6)	9	38.37	29.9
Preg (Ec. Polytechnique) (33)(-3)	10	36.41	28.4
Centre (Ens Cachan) (7)(-3)	11	34.33	26.8
Centre *(Cired)* (14)(+6)	12	33.41	26.1
Ces (Pse-Paris 1) (138)(-2)	13	29.13	22.7
Smart (Inra Rennes) (12)(-1)	14	28.59	22.3
Lameta *(Montpellier 1-Inra)* (33)(+16)	15	22.34	17.4
Thema (Cergy Pontoise) (36)(-6)	16	22.19	17.3
Centre (Strasbourg 3) (13)(0)	17	22.02	17.2
Cerdi *(Clermont 1)* (26)(+12)	18	20.63	16.1
Gate *(Lyon 2)* (27)(+4)	19	19.52	15.2
Gael *(Grenoble 2-Inra)* (15)(+5)	20	18.86	14.7
Economix (Paris 10) (66)(-1)	21	18.11	14.1
Centre *(Lille 2)* (13)(+20)	22	17.96	14.0
Iredu-Eco *(Dijon)* (12)(+59)	23	17.71	13.8
Non-Cermes *(Paris 11)* (16)(+31)	24	17.59	13.7
Ermes (Paris 2) (17)(+1)	25	17.55	13.7
Centre (Besancon) (24)(+1)	26	16.76	13.1
Beta (Nancy 2-Strasb. 1) (71)(+1)	27	16.70	13.0
Autre (Tse-Toulouse 1) (23)(-6)	28	16.38	12.8
Drm *(Paris 9)* (69)(+45)	29	15.97	12.5
Centre (La Rochelle) (5)(-14)	30	13.98	10.9
Centre *(Chambery)* (15)(+28)	31	13.79	10.8
Cepn *(Paris 13)* (33)(+16)	32	13.50	10.5
Centre *(Vers. St Quentin)* (24)(+34)	33	13.32	10.4
Lef *(Inra Nancy)* (7)(+18)	34	12.92	10.1
Centre (Nantes) (23)(-15)	35	12.81	10.0
Lirhe-Eco *(Tse-Toulouse 1)* (20)(+54)	36	12.62	9.9
Cermes *(Paris 11)* (21)(+64)	37	12.36	9.6
Centre *(Pau)* (17)(+15)	38	12.26	9.6
Gretha (Bordeaux 4) (37)(-5)	39	11.90	9.3
Centre *(Paris 7)* (10)(+45)	40	11.47	8.9
Cesaer (Inra Dijon) (11)(-22)	41	11.45	8.9
Leg *(Dijon)* (42)(+37)	42	11.43	8.9
Centre (Le Mans) (18)(-29)	43	11.20	8.7
Clerse-Eco *(Lille 1-Poly. Lille)* (36)(+14)	44	9.93	7.7
Lest-Eco *(Aix Marseille 2-3)* (13)(+37)	45	9.82	7.7
Aliss (Inra Ivry) (19)(-22)	46	9.57	7.5
Centre (Evry) (18)(-15)	47	9.41	7.3
Mona-Tsv *(Inra Ivry)* (19)(+54)	48	9.20	7.2
Gredeg *(Nice)* (71)(+17)	49	9.00	7.0
Moisa (Montpellier 1-Inra) (8)(-4)	50	8.62	6.7
Non-Drm (Paris 9) (56)(-6)	51	8.22	6.4
Lepii *(Grenoble 2-Inra)* (29)(+38)	52	8.10	6.3
Non-Gretha *(Bordeaux 4)* (35)(+9)	53	8.08	6.3
Centre (Lille 3) (11)(-21)	54	8.03	6.3
Non-Gate *(Lyon 2)* (43)(+25)	55	7.97	6.2
Leo (Orleans) (30)(-13)	56	7.85	6.1
Centre *(Rennes 2)* (7)(+7)	57	7.73	6.0
Cerag *(Grenoble 2-Inra)* (48)(+34)	58	7.48	5.8
Centre (Montpellier 3) (10)(-23)	59	7.43	5.8
Centre (Perpignan) (12)(-31)	60	7.41	5.8

suite page suivante

suite de la page précédente

centre	rg.	tot.	nor.
Lest-Eco *(Aix Marseille 2-3)* (13)(+29)	61	127.7	1.7
Cesaer (Inra Dijon) (11)(-17)	62	125.9	1.7
Autre (Grenoble 2-Inra) (38)(-20)	63	120.2	1.6
Centre *(Paris 7)* (10)(+31)	64	114.7	1.5
Autre (Montpellier 1-Inra) (21)(-16)	65	108.4	1.4
Non-Crem (Caen-Rennes 1) (39)(-8)	66	108.3	1.4
Centre *(Rouen)* (22)(+19)	67	106.9	1.4
Centre (Limoges) (18)(-9)	68	96.5	1.3
Centre (Lille 3) (11)(-14)	69	88.4	1.2
Centre (La Reunion) (19)(-19)	70	87.4	1.2
Centre (Perpignan) (12)(-23)	71	85.3	1.1
Lef *(Inra Nancy)* (7)(+15)	72	84.0	1.1
Non-Beta *(Nancy 2-Strasb. 1)* (24)(+4)	73	77.1	1.0
Non-Cepn *(Paris 13)* (12)(+4)	74	75.4	1.0
Centre *(Angers)* (18)(+5)	75	73.3	1.0
Centre (Montpellier 3) (10)(-6)	76	70.6	0.9
Centre (La Rochelle) (5)(-8)	77	69.9	0.9
Moisa (Montpellier 1-Inra) (8)(+1)	78	69.0	0.9
Non-Gredeg (Nice) (12)(-6)	79	63.4	0.8
Centre (Toulon) (11)(-5)	80	63.2	0.8
Centre (Poitiers) (27)(-29)	81	61.3	0.8
Centre (Tours) (12)(-21)	82	59.2	0.8
Autre *(Dijon)* (11)(+14)	83	58.9	0.8
Centre (Littoral) (13)(-22)	84	54.2	0.7
Centre *(Rennes 2)* (7)(+4)	85	54.1	0.7
Centre (Marne La Vallee) (11)(-19)	86	53.3	0.7
Centre (Mulhouse) (9)(-15)	87	52.4	0.7
Centre *(Toulouse 2)* (10)(+15)	88	52.2	0.7
Centre (Brest) (21)(-25)	89	51.2	0.7
Centre (Ant. Guy.) (22)(-25)	90	50.5	0.7
Centre *(Valenciennes)* (8)(+3)	91	48.0	0.6
Centre (Le Havre) (15)(-4)	92	42.5	0.6
Centre (Cnam) (7)(-11)	93	42.4	0.6
Centre (Artois) (13)(-13)	94	39.8	0.5
Centre (Metz) (13)(-11)	95	38.6	0.5
Centre (Paris 5) (11)(-20)	96	33.1	0.4
Centre (Amiens) (22)(-14)	97	29.7	0.4
Centre (Toulouse 3) (10)(+2)	98	27.7	0.4
Non-Economix (Paris 10) (15)(-6)	99	26.4	0.4
Centre (Lyon 1) (5)(-4)	100	24.5	0.3
Centre (Corte) (9)(-3)	101	15.9	0.2
Centre (Bretagne Sud) (9)(-11)	102	13.3	0.2
Centre (Lyon 3) (9)(-1)	103	8.1	0.1
	104		

centre	rg.	p.c.	nor.
Crem (Caen-Rennes 1) (82)(-21)	61	7.20	5.6
Centre (Paris 8) (27)(-13)	62	6.87	5.4
Non-Ermes (Paris 2) (24)(-22)	63	6.76	5.3
Centre (St Etienne) (22)(-10)	64	6.61	5.2
Centre *(Valenciennes)* (8)(+10)	65	6.40	5.0
Non-Cepn (Paris 13) (12)(+2)	66	6.28	4.9
Centre (Reims) (32)(-5)	67	6.27	4.9
Centre (Mulhouse) (9)(-33)	68	6.16	4.8
Centre (Cnam) (7)(-22)	69	6.05	4.7
Autre *(Pse-Paris 1)* (33)(+19)	70	5.81	4.5
Centre (Toulon) (11)(-14)	71	5.75	4.5
Autre (Lille 1-Poly. Lille) (31)(-35)	72	5.68	4.4
Centre *(Toulouse 2)* (10)(+30)	73	5.49	4.3
Autre *(Dijon)* (11)(+20)	74	5.36	4.2
Centre (Limoges) (18)(-13)	75	5.36	4.2
Non-Gredeg (Nice) (12)(-16)	76	5.29	4.1
Autre (Montpellier 1-Inra) (21)(-26)	77	5.16	4.0
Centre (Marne La Vallee) (11)(-40)	78	5.08	4.0
Lem (Lille 1-Poly. Lille) (87)(-1)	79	5.02	3.9
Centre (Tours) (12)(-41)	80	4.93	3.9
Centre (Lyon 1) (5)(-16)	81	4.91	3.8
Centre *(Rouen)* (22)(+13)	82	4.86	3.8
Centre *(Paris 12)* (30)(+3)	83	4.73	3.7
Centre (La Reunion) (19)(-34)	84	4.60	3.6
Centre (Littoral) (13)(-41)	85	4.17	3.3
Centre (Angers) (18)(-5)	86	4.07	3.2
Autre *(Aix Marseille 2-3)* (57)(+11)	87	3.58	2.8
Non-Beta (Nancy 2-Strasb. 1) (24)(-1)	88	3.28	2.6
Autre (Grenoble 2-Inra) (38)(-19)	89	3.21	2.5
Centre (Artois) (13)(-18)	90	3.19	2.5
Centre (Paris 5) (11)(-35)	91	3.16	2.5
Centre (Metz) (13)(-16)	92	3.09	2.4
Centre (Le Havre) (15)(-9)	93	2.83	2.2
Non-Crem (Caen-Rennes 1) (39)(-6)	94	2.78	2.2
Centre *(Toulouse 3)* (10)(+4)	95	2.77	2.2
Centre (Brest) (21)(-27)	96	2.50	2.0
Centre (Ant. Guy.) (22)(-24)	97	2.35	1.8
Centre (Poitiers) (27)(-27)	98	2.27	1.8
Non-Economix (Paris 10) (15)(-3)	99	1.82	1.4
Centre (Corte) (9)(-4)	100	1.77	1.4
Centre (Bretagne Sud) (9)(-24)	101	1.48	1.2
Centre (Amiens) (22)(-9)	102	1.35	1.1
Centre (Lyon 3) (9)(-3)	103	0.90	0.7
Non-Cerdi (Clermont 1) (6)(0)	104	0.01	0

La colonne "rg." donne le rang, la colonne "tot." donne le score total, "nor." le score normalisé par rapport à celui du premier classé, "p.c." donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.4.2.1. Indice G

TABLE 1.23 – Centres, Indice G, T=Dégressif

centre	rg.	tot.	nor.
Pjse (Pse-Paris 1) (43)(+2)	1	96.2	100.0
Gremaq (Tse-Toulouse 1) (59)(-1)	2	92.5	96.2
Ces (Pse-Paris 1) (138)(-1)	3	76.5	79.6
Greghec (Hec) (75)(+1)	4	72.7	75.6

centre	rg.	p.c.	nor.
Centre *(Iep Paris)* (9)(+4)	1	9.23	100.0
Pjse (Pse-Paris 1) (43)(0)	2	9.21	99.8
Lerna (Tse-Toulouse 1) (19)(0)	3	7.10	76.9
Gremaq (Tse-Toulouse 1) (59)(-3)	4	6.59	71.3

suite page suivante

suite de la page précédente

centre	**rg.**	**tot.**	**nor.**
Grecsta (Crest-Ensae) (31)(+1)	5	68.7	71.5
Non-Grecsta *(Crest-Ensae)* (36)(+6)	6	63.1	65.7
Greqam (Aix Marseille 2-3) (45)(-3)	7	60.8	63.3
Beta (Nancy 2-Strasb. 1) (71)(+2)	8	51.5	53.6
Preg (Ec. Polytechnique) (33)(0)	9	45.6	47.5
Economix (Paris 10) (66)(-2)	10	44.8	46.6
Lerna (Tse-Toulouse 1) (19)(-4)	11	43.7	45.5
Centre *(Iep Paris)* (9)(+6)	12	43.6	45.3
Drm *(Paris 9)* (69)(+14)	13	42.3	43.9
Lameta (Montpellier 1-Inra) (33)(+2)	14	39.7	41.3
Gate (Lyon 2) (27)(0)	15	35.4	36.8
Gredeg (Nice) (71)(+1)	16	34.3	35.7
Eco. Pub. *(Inra Vers-Grig)* (12)(+9)	17	34.0	35.3
Thema (Cergy Pontoise) (36)(-7)	18	32.1	33.4
Ermes *(Paris 2)* (17)(+14)	19	31.1	32.4
Centre (Le Mans) (18)(-1)	20	31.0	32.2
Centre *(Cired)* (14)(+16)	21	30.7	32.0
Cerdi *(Clermont 1)* (26)(+3)	22	30.0	31.2
Lem (Lille 1-Poly. Lille) (87)(0)	23	29.3	30.5
Crem (Caen-Rennes 1) (82)(-11)	24	29.0	30.2
Gael *(Grenoble 2-Inra)* (15)(+15)	25	27.4	28.5
Centre *(Strasbourg 3)* (13)(+9)	26	27.2	28.3
Lirhe-Eco *(Tse-Toulouse 1)* (20)(+59)	27	27.2	28.3
Leg *(Dijon)* (42)(+19)	28	26.5	27.5
Smart (Inra Rennes) (12)(+1)	29	26.2	27.3
Centre *(Ens Cachan)* (7)(+8)	30	25.8	26.8
Centre (Besancon) (24)(-7)	31	24.5	25.5
Non-Cermes *(Paris 11)* (16)(+31)	32	24.4	25.4
Centre *(Vers. St Quentin)* (24)(+20)	33	24.0	25.0
Non-Drm (Paris 9) (56)(-20)	34	23.9	24.8
Cepn (Paris 13) (33)(-4)	35	23.8	24.7
Centre *(Lille 2)* (13)(+24)	36	23.7	24.7
Autre (Tse-Toulouse 1) (23)(-16)	37	23.3	24.2
Gretha (Bordeaux 4) (37)(-18)	38	23.2	24.1
Centre *(Chambery)* (15)(+27)	39	22.8	23.7
Autre *(Pse-Paris 1)* (33)(+28)	40	22.7	23.6
Centre (Nantes) (23)(-19)	41	22.2	23.1
Non-Gate *(Lyon 2)* (43)(+5)	42	22.2	23.1
Non-Gretha (Bordeaux 4) (35)(-2)	43	21.6	22.5
Cerag *(Grenoble 2-Inra)* (48)(+10)	44	21.5	22.4
Autre *(Aix Marseille 2-3)* (57)(+12)	45	21.0	21.8
Centre *(Pau)* (17)(+10)	46	20.9	21.7
Clerse-Eco (Lille 1-Poly. Lille) (36)(-12)	47	20.9	21.7
Centre (Reims) (32)(-4)	48	20.3	21.1
Aliss (Inra Ivry) (19)(-20)	49	19.1	19.9
Centre (Evry) (18)(-14)	50	18.7	19.4
Cermes *(Paris 11)* (21)(+48)	51	18.5	19.2
Iredu-Eco *(Dijon)* (12)(+40)	52	18.4	19.1
Non-Ermes (Paris 2) (24)(-10)	53	18.4	19.2
Leo (Orleans) (30)(-22)	54	18.1	18.8
Autre (Lille 1-Poly. Lille) (31)(-27)	55	17.5	18.2
Centre *(Limoges)* (18)(+3)	56	17.2	17.9
Cesaer (Inra Dijon) (11)(-12)	57	17.1	17.8
Mona-Tsv *(Inra Ivry)* (19)(+43)	58	16.8	17.5
Non-Crem (Caen-Rennes 1) (39)(-1)	59	16.7	17.3
Lepii *(Grenoble 2-Inra)* (29)(+14)	60	16.4	17.0
Centre *(Paris 7)* (10)(+34)	61	16.3	17.0
Centre (St Etienne) (22)(-12)	62	15.9	16.5
Centre *(Paris 12)* (30)(+8)	63	15.3	15.9
Centre *(Angers)* (18)(+16)	64	15.2	15.8
Lest-Eco *(Aix Marseille 2-3)* (13)(+25)	65	14.6	15.2
Autre (Montpellier 1-Inra) (21)(-17)	66	14.4	14.9
Centre (Lille 3) (11)(-11)	67	14.4	15.0
Lef *(Inra Nancy)* (7)(+19)	68	13.8	14.3

centre	**rg.**	**p.c.**	**nor.**
Centre *(Cired)* (14)(+13)	5	5.53	59.9
Eco. Pub. *(Inra Vers-Grig)* (12)(+3)	6	5.34	57.9
Greqam (Aix Marseille 2-3) (45)(-3)	7	5.18	56.1
Smart *(Inra Rennes)* (12)(+5)	8	4.95	53.6
Preg (Ec. Polytechnique) (33)(-2)	9	4.83	52.3
Grecsta (Crest-Ensae) (31)(-4)	10	4.78	51.8
Centre (Ens Cachan) (7)(-3)	11	4.35	47.1
Iredu-Eco *(Dijon)* (12)(+70)	12	4.02	43.6
Non-Grecsta (Crest-Ensae) (36)(-1)	13	3.90	42.2
Ces (Pse-Paris 1) (138)(-3)	14	3.87	42.0
Cerdi *(Clermont 1)* (26)(+15)	15	3.64	39.4
Greghec (Hec) (75)(-1)	16	3.51	38.0
Thema (Cergy Pontoise) (36)(-7)	17	3.47	37.6
Gate *(Lyon 2)* (27)(+5)	18	3.46	37.4
Lef *(Inra Nancy)* (7)(+34)	19	3.46	37.5
Centre (Strasbourg 3) (13)(-3)	20	3.24	35.1
Economix (Paris 10) (66)(0)	21	3.24	35.1
Centre *(Chambery)* (15)(+37)	22	3.22	34.9
Cepn *(Paris 13)* (33)(+25)	23	3.02	32.7
Cermes *(Paris 11)* (21)(+77)	24	2.90	31.4
Gael (Grenoble 2-Inra) (15)(0)	25	2.85	30.9
Beta (Nancy 2-Strasb. 1) (71)(+2)	26	2.83	30.7
Lest-Eco *(Aix Marseille 2-3)* (13)(+55)	27	2.75	29.8
Moisa *(Montpellier 1-Inra)* (8)(+18)	28	2.68	29.0
Lameta (Montpellier 1-Inra) (33)(+2)	29	2.56	27.7
Gretha *(Bordeaux 4)* (37)(+4)	30	2.52	27.3
Lirhe-Eco *(Tse-Toulouse 1)* (20)(+59)	31	2.51	27.2
Centre (La Rochelle) (5)(-16)	32	2.48	26.9
Centre (Perpignan) (12)(-4)	33	2.43	26.3
Centre (Besancon) (24)(-7)	34	2.39	25.8
Centre *(Vers. St Quentin)* (24)(+32)	35	2.36	25.6
Centre (Le Mans) (18)(-22)	36	2.33	25.3
Ermes (Paris 2) (17)(-10)	37	2.33	25.3
Centre (Montpellier 3) (10)(-2)	38	2.32	25.2
Non-Cermes *(Paris 11)* (16)(+17)	39	2.32	25.1
Centre (Nantes) (23)(-20)	40	2.30	24.9
Centre *(Rennes 2)* (7)(+23)	41	2.27	24.6
Centre (Evry) (18)(-10)	42	2.21	23.9
Cesaer (Inra Dijon) (11)(-24)	43	2.20	23.8
Lepii *(Grenoble 2-Inra)* (29)(+46)	44	2.18	23.6
Centre *(Pau)* (17)(+8)	45	2.17	23.5
Crem (Caen-Rennes 1) (82)(-5)	46	2.17	23.5
Drm *(Paris 9)* (69)(+27)	47	2.15	23.3
Autre (Tse-Toulouse 1) (23)(-26)	48	2.12	23.0
Centre (Lille 3) (11)(-15)	49	2.12	23.0
Aliss (Inra Ivry) (19)(-26)	50	2.08	22.6
Clerse-Eco *(Lille 1-Poly. Lille)* (36)(+7)	51	2.02	21.8
Non-Cepn *(Paris 13)* (12)(+16)	52	1.99	21.6
Mona-Tsv *(Inra Ivry)* (19)(+49)	53	1.93	20.9
Centre (Cnam) (7)(-7)	54	1.92	20.8
Leg *(Dijon)* (42)(+25)	55	1.92	20.8
Gredeg *(Nice)* (71)(+10)	56	1.86	20.2
Leo (Orleans) (30)(-14)	57	1.85	20.0
Non-Drm (Paris 9) (56)(-13)	58	1.84	19.9
Cerag *(Grenoble 2-Inra)* (48)(+33)	59	1.82	19.8
Centre *(Paris 7)* (10)(+25)	60	1.74	18.8
Centre (St Etienne) (22)(-7)	61	1.73	18.7
Centre (Reims) (32)(0)	62	1.68	18.2
Centre (Lille 2) (13)(-21)	63	1.67	18.1
Non-Gate *(Lyon 2)* (43)(+16)	64	1.66	18.0
Centre (Paris 8) (27)(-16)	65	1.62	17.5
Centre (Toulon) (11)(-8)	66	1.62	17.5
Non-Gredeg (Nice) (12)(-7)	67	1.60	17.3
Centre (Marne La Vallee) (11)(-30)	68	1.56	16.9

suite page suivante

suite de la page précédente

centre	**rg.**	**tot.**	**nor.**
Centre (Paris 8) (27)(-30)	69	13.5	14.0
Centre (Rouen) (22)(+16)	70	13.2	13.8
Centre (La Reunion) (19)(-20)	71	13.1	13.6
Centre (Valenciennes) (8)(+22)	72	13.0	13.6
Autre (Grenoble 2-Inra) (38)(-30)	73	12.9	13.4
Centre (Montpellier 3) (10)(-3)	74	12.9	13.4
Moisa (Montpellier 1-Inra) (8)(+4)	75	12.8	13.3
Non-Beta (Nancy 2-Strasb. 1) (24)(+2)	76	12.8	13.3
Centre (Perpignan) (12)(-29)	77	12.5	13.0
Centre (Toulon) (11)(-3)	78	12.2	12.6
Centre (Mulhouse) (9)(-7)	79	11.9	12.4
Centre (Poitiers) (27)(-28)	80	11.8	12.3
Centre (Tours) (12)(-20)	81	11.7	12.2
Centre (Artois) (13)(-1)	82	11.1	11.5
Centre (Brest) (21)(-18)	83	11.1	11.5
Autre (Dijon) (11)(+13)	84	10.6	11.0
Centre (Marne La Vallee) (11)(-18)	85	10.4	10.8
Centre (Toulouse 2) (10)(+18)	86	10.4	10.8
Non-Gredeg (Nice) (12)(-14)	87	10.3	10.7
Centre (Le Havre) (15)(0)	88	10.1	10.5
Non-Cepn (Paris 13) (12)(-11)	89	9.9	10.3
Centre (Metz) (13)(-6)	90	9.7	10.1
Centre (Ant. Guy.) (22)(-26)	91	9.6	10.0
Centre (Cnam) (7)(-10)	92	9.4	9.8
Centre (La Rochelle) (5)(-24)	93	9.3	9.7
Centre (Rennes 2) (7)(-4)	94	9.3	9.7
Centre (Paris 5) (11)(-19)	95	8.1	8.4
Non-Economix (Paris 10) (15)(-3)	96	7.8	8.1
Centre (Littoral) (13)(-35)	97	7.6	7.9
Centre (Toulouse 3) (10)(+2)	98	5.9	6.1
Centre (Corte) (9)(-1)	99	5.2	5.4
Centre (Lyon 3) (9)(+3)	100	5.2	5.4
Centre (Lyon 1) (5)(-5)	101	4.6	4.8
Centre (Amiens) (22)(-19)	102	4.5	4.7
Centre (Bretagne Sud) (9)(-12)	103	4.2	4.4
	104		

centre	**rg.**	**p.c.**	**nor.**
Non-Gretha (Bordeaux 4) (35)(-7)	69	1.51	16.4
Centre (Mulhouse) (9)(-35)	70	1.49	16.2
Centre (Toulouse 2) (10)(+33)	71	1.49	16.2
Autre (Pse-Paris 1) (33)(+17)	72	1.45	15.7
Centre (Rouen) (22)(+22)	73	1.42	15.4
Non-Ermes (Paris 2) (24)(-33)	74	1.39	15.0
Autre (Montpellier 1-Inra) (21)(-24)	75	1.38	14.9
Centre (La Reunion) (19)(-26)	76	1.36	14.8
Lem (Lille 1-Poly. Lille) (87)(+2)	77	1.36	14.7
Autre (Lille 1-Poly. Lille) (31)(-41)	78	1.35	14.6
Centre (Littoral) (13)(-35)	79	1.34	14.5
Centre (Paris 12) (30)(+6)	80	1.29	14.0
Centre (Le Havre) (15)(+3)	81	1.28	13.8
Centre (Tours) (12)(-43)	82	1.21	13.1
Autre (Grenoble 2-Inra) (38)(-13)	83	1.20	13.0
Centre (Limoges) (18)(-23)	84	1.19	12.8
Autre (Dijon) (11)(+9)	85	1.01	11.0
Centre (Toulouse 3) (10)(+14)	86	1.01	10.9
Centre (Valenciennes) (8)(-10)	87	1.01	11.0
Autre (Aix Marseille 2-3) (57)(+10)	88	1.00	10.8
Centre (Paris 5) (11)(-33)	89	0.93	10.1
Centre (Lyon 1) (5)(-25)	90	0.92	9.9
Non-Beta (Nancy 2-Strasb. 1) (24)(-4)	91	0.91	9.9
Centre (Metz) (13)(-16)	92	0.88	9.5
Centre (Poitiers) (27)(-21)	93	0.88	9.5
Centre (Brest) (21)(-25)	94	0.83	9.0
Centre (Corte) (9)(+2)	95	0.83	9.0
Centre (Ant. Guy.) (22)(-23)	96	0.81	8.8
Non-Crem (Caen-Rennes 1) (39)(-8)	97	0.81	8.7
Centre (Angers) (18)(-17)	98	0.78	8.5
Centre (Bretagne Sud) (9)(-22)	99	0.77	8.3
Centre (Artois) (13)(-28)	100	0.75	8.1
Centre (Amiens) (22)(-8)	101	0.63	6.8
Non-Economix (Paris 10) (15)(-6)	102	0.52	5.6
Centre (Lyon 3) (9)(-3)	103	0.45	4.9
Non-Cerdi (Clermont 1) (6)(0)	104	0.01	0.1

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.5. Variantes

1.5.1. Classements par année de carrière et sur les moins de 50 ans

Les universités et centres de recherche français présentent des structures par âge relativement différentes. On peut se demander si celles-ci peuvent avoir une influence sur les classements obtenus et ce qu'il adviendrait de ceux-ci si l'âge des chercheurs étaient pris en compte. A cette fin, nous présentons tout d'abord en annexe des classements établis pour notre période de temps T=Carrière, qui ramène le score de chaque chercheur à son âge (moins 26 ans). De plus, une fraction importante de la population a un âge compris entre 55 et 65 ans. Parmi les chercheurs de cette tranche d'âge, il se trouve un certain nombre de personnes avec des stocks d'articles assez important et qui contribuent de manière substantielle au capital recherche de leur centre/université. Certains étant proches de la retraite, il est intéressant de voir où se situent les capitaux recherche des centres en l'absence des chercheurs les plus âgés, ce qui pourrait être la situation du centre ou de l'université dans quelques années. Si le rang d'un centre diminue par rapport au classement avec tous les membres cela peut indiquer, par exemple, que des recrutements sont à envisager si le centre veut maintenir sa position. Une coupure, relativement arbitraire, à 50 ans est effectuée, seuls les chercheurs dont l'âge est strictement inférieur à 50 ans étant retenus pour le calcul des scores des institutions également présentés en annexe.

1.5.2. Classements ne gardant que les 10 ou 30 chercheurs les plus productifs

La comparaison des centres ou universités est parfois rendue difficile par les différences de taille des institutions, raison pour laquelle nous présentons à la fois des classements en volume et par chercheur donnant deux éclairages complémentaires. Pour contrôler l'effet taille (le volume de citations augmente avec le nombre de chercheurs mais les grandes institutions ont en général aussi une plus grande part de chercheurs peu productifs ce qui diminue leur productivité moyenne), nous présentons dans cette section une approche différente qui consiste à uniformiser le nombre de chercheurs de toutes les institutions en ne retenant que les 10 puis 30 chercheurs les plus productifs de chacune. Les grandes institutions gardent un certain avantage lié à leur taille, puisque la distribution de productivité de leur chercheur est alors probablement tronquée à un niveau plus élevé mais l'effet n'est cependant plus totalement mécanique. Production totale et par chercheur conduisent d'ailleurs à la même hiérarchie des institutions, hormis du fait des quelques institutions n'ayant pas au moins 10 (respectivement 30) chercheurs. Les tableaux donnant ces classements sont également donné en annexe. Là encore, si les rangs de certaines institutions varient légèrement, la hiérarchie globale est relativement maintenue, en tous les cas dans le haut du classement. Les variations deviennent plus importantes lorsqu'on les descend, ce qui peut donner des indications utiles quant à la distribution de la production à l'intérieur des institutions.

1.6. Conclusions

Le résumé des résultats que nous obtenons est donné à la fin de l'introduction de cet article. La conclusion principale est que l'approche GS ne modifie pas fondamentalement l'image obtenue de la recherche française en économie et la hiérarchie en termes de publications académiques des centres de recherche et universités français. Pour certaines institutions dont l'activité n'est pas au cœur de l'économie, les évolutions sont un peu plus sensibles, ce qui peut faire de GS un instrument complémentaire intéressant d'Econlit, en espérant que ses limites techniques et erreurs de mesure soient rapidement comblées.

Une fois l'outil de mesure défini et ses propriétés étudiées et comprises, une étape importante, mais cependant complexe, reste à effectuer. Elle consiste en l'étude des déterminants de l'output en termes de publications des différents centres de recherche ou universités français. Il s'agit de la tâche que nous essayons de mener dans Bosquet et

Combes (2012). Nous expliquons successivement le fait qu'un chercheur publie, la quantité qu'il publie et la qualité de ses publications à la fois par un certain nombre de variables individuelles, trouvées classiquement en économie du travail comme le sexe et l'âge, et par des effets locaux liés aux caractéristiques des institutions auxquelles il est affilié. Les effets qui sont plus particulièrement étudiés sont ceux de la taille totale (nombre de chercheurs) de l'institution et de ses spécialisation et diversité thématiques, qui peuvent être appréhendés facilement et précisément en économie via les codes JEL présents dans la base de données Econlit. Nous contrôlons pour l'importance de l'activité d'enseignement et de recherche via le nombre d'étudiants formés par chercheur. Le rôle de l'hétérogénéité des centres en termes de productivité de leurs chercheurs, et la présence ou non de "stars", est un autre exemple d'élément pouvant affecter la productivité locale des chercheurs que nous étudions. Une fois le diagnostic de diversité des profils de publication posé, passer à l'étape de quantification des déterminants de cette diversité est certainement primordial en termes de politiques d'éducation et de recherche.

APPENDIX

1.A. Classements par année de carrière et sur les moins de 50 ans

1.A.1. Centres, T=Carrière

TABLE 1.24 – Centres, Citations totales, T=Carrière

centre	rg.	tot.	nor.
Gremaq (Tse-Toulouse 1) (59)(0)	1	2101.1	100.0
Pjse (Pse-Paris 1) (43)(0)	2	1347.7	64.1
Ces (Pse-Paris 1) (138)(0)	3	897.5	42.7
Greghec (Hec) (75)(+2)	4	756.3	36.0
Grecsta (Crest-Ensae) (31)(0)	5	539.8	25.7
Greqam (Aix Marseille 2-3) (45)(-2)	6	444.4	21.2
Preg (Ec. Polytechnique) (33)(+1)	7	337.0	16.0
Lerna (Tse-Toulouse 1) (19)(+1)	8	321.0	15.3
Economix (Paris 10) (66)(-2)	9	319.3	15.2
Centre *(Iep Paris)* (9)(+5)	10	285.7	13.6
Non-Grecsta (Crest-Ensae) (36)(+2)	11	282.7	13.5
Beta (Nancy 2-Strasb. 1) (71)(-2)	12	269.2	12.8
Thema (Cergy Pontoise) (36)(-2)	13	224.4	10.7
Drm *(Paris 9)* (69)(+20)	14	221.5	10.5
Crem (Caen-Rennes 1) (82)(-3)	15	168.6	8.0
Gredeg *(Nice)* (71)(+9)	16	161.3	7.7
Non-Drm (Paris 9) (56)(-3)	17	149.0	7.1
Eco. Pub. *(Inra Vers-Grig)* (12)(+4)	18	139.8	6.7
Cerag *(Grenoble 2-Inra)* (48)(+40)	19	136.8	6.5
Cerdi (Clermont 1) (26)(-3)	20	131.6	6.3
Gate (Lyon 2) (27)(0)	21	129.8	6.2
Centre (Besancon) (24)(-2)	22	123.2	5.9
Leg *(Dijon)* (42)(+31)	23	119.2	5.7
Lem *(Lille 1-Poly. Lille)* (87)(+3)	24	110.6	5.3
Cepn *(Paris 13)* (33)(+15)	25	109.0	5.2
Non-Gate *(Lyon 2)* (43)(+15)	26	103.9	4.9
Gretha (Bordeaux 4) (37)(-8)	27	100.9	4.8
Lameta (Montpellier 1-Inra) (33)(-4)	28	97.0	4.6
Centre *(Cired)* (14)(+14)	29	94.7	4.5
Centre *(Vers. St Quentin)* (24)(+22)	30	91.9	4.4
Centre (Le Mans) (18)(-13)	31	88.6	4.2
Smart (Inra Rennes) (12)(-6)	32	87.9	4.2
Clerse-Eco *(Lille 1-Poly. Lille)* (36)(+15)	33	87.8	4.2
Centre (Nantes) (23)(-11)	34	81.5	3.9
Autre (Tse-Toulouse 1) (23)(-7)	35	72.0	3.4
Ermes (Paris 2) (17)(-3)	36	71.2	3.4
Leo (Orleans) (30)(-21)	37	65.6	3.1
Non-Cermes *(Paris 11)* (16)(+29)	38	64.0	3.1
Centre (Strasbourg 3) (13)(-7)	39	63.7	3.0
Non-Gretha *(Bordeaux 4)* (35)(+4)	40	62.3	3.0
Autre *(Aix Marseille 2-3)* (57)(+19)	41	60.5	2.9
Centre *(Lille 2)* (13)(+8)	42	60.4	2.9
Cermes *(Paris 11)* (21)(+58)	43	55.9	2.7
Centre (Evry) (18)(-13)	44	55.0	2.6
Gael (Grenoble 2-Inra) (15)(+1)	45	55.0	2.6
Centre (Ens Cachan) (7)(-9)	46	54.6	2.6

centre	rg.	p.c.	nor.
Gremaq (Tse-Toulouse 1) (59)(+1)	1	35.49	100.0
Centre (Iep Paris) (9)(+2)	2	33.62	94.7
Pjse (Pse-Paris 1) (43)(-2)	3	31.23	88.0
Lerna (Tse-Toulouse 1) (19)(-1)	4	17.35	48.9
Grecsta (Crest-Ensae) (31)(0)	5	17.30	48.8
Eco. Pub. (Inra Vers-Grig) (12)(+2)	6	11.65	32.8
Greghec *(Hec)* (75)(+9)	7	10.10	28.5
Preg (Ec. Polytechnique) (33)(-1)	8	10.09	28.4
Greqam (Aix Marseille 2-3) (45)(-3)	9	9.91	27.9
Non-Grecsta *(Crest-Ensae)* (36)(+4)	10	7.87	22.2
Centre (Ens Cachan) (7)(-2)	11	7.45	21.0
Smart (Inra Rennes) (12)(-2)	12	7.33	20.6
Centre *(Cired)* (14)(+19)	13	7.02	19.8
Ces (Pse-Paris 1) (138)(-2)	14	6.52	18.4
Thema (Cergy Pontoise) (36)(-2)	15	6.32	17.8
Lef *(Inra Nancy)* (7)(+31)	16	5.29	14.9
Centre (Besancon) (24)(+2)	17	5.13	14.5
Cerdi *(Clermont 1)* (26)(+3)	18	5.06	14.3
Centre (Strasbourg 3) (13)(-4)	19	4.98	14.0
Economix (Paris 10) (66)(-2)	20	4.87	13.7
Centre (Le Mans) (18)(-10)	21	4.86	13.7
Gate *(Lyon 2)* (27)(+7)	22	4.74	13.4
Centre *(Lille 2)* (13)(+12)	23	4.65	13.1
Ermes *(Paris 2)* (17)(+4)	24	4.19	11.8
Non-Cermes *(Paris 11)* (16)(+32)	25	4.13	11.6
Iredu-Eco *(Dijon)* (12)(+51)	26	4.03	11.4
Centre *(Vers. St Quentin)* (24)(+28)	27	3.83	10.8
Beta *(Nancy 2-Strasb. 1)* (71)(+3)	28	3.79	10.7
Gael *(Grenoble 2-Inra)* (15)(+8)	29	3.79	10.7
Centre (Nantes) (23)(-8)	30	3.62	10.2
Cepn *(Paris 13)* (33)(+27)	31	3.35	9.5
Drm *(Paris 9)* (69)(+52)	32	3.23	9.1
Centre *(Chambery)* (15)(+21)	33	3.17	8.9
Autre (Tse-Toulouse 1) (23)(0)	34	3.15	8.9
Centre (Mulhouse) (9)(-5)	35	3.12	8.8
Centre (Evry) (18)(-12)	36	3.06	8.6
Centre *(Paris 7)* (10)(+48)	37	2.95	8.3
Lameta (Montpellier 1-Inra) (33)(0)	38	2.94	8.3
Cerag *(Grenoble 2-Inra)* (48)(+53)	39	2.88	8.1
Leg *(Dijon)* (42)(+43)	40	2.85	8.0
Gretha (Bordeaux 4) (37)(-4)	41	2.77	7.8
Cesaer (Inra Dijon) (11)(-22)	42	2.76	7.8
Cermes *(Paris 11)* (21)(+60)	43	2.73	7.7
Non-Drm (Paris 9) (56)(-4)	44	2.66	7.5
Lirhe-Eco *(Tse-Toulouse 1)* (20)(+51)	45	2.61	7.4
Clerse-Eco *(Lille 1-Poly. Lille)* (36)(+24)	46	2.47	7.0

suite page suivante

suite de la page précédente

centre	rg.	tot.	nor.
Lirhe-Eco *(Tse-Toulouse 1)* (20)(+42)	47	52.3	2.5
Centre (Reims) (32)(+1)	48	51.1	2.4
Autre *(Pse-Paris 1)* (33)(+15)	49	49.3	2.3
Iredu-Eco *(Dijon)* (12)(+36)	50	48.4	2.3
Centre *(Chambery)* (15)(+15)	51	47.6	2.3
Centre (St Etienne) (22)(-6)	52	47.5	2.3
Lepii *(Grenoble 2-Inra)* (29)(+23)	53	47.1	2.2
Centre (Paris 8) (27)(-15)	54	43.7	2.1
Autre (Lille 1-Poly. Lille) (31)(-26)	55	42.6	2.0
Non-Ermes (Paris 2) (24)(-20)	56	41.8	2.0
Centre *(Rouen)* (22)(+21)	57	41.1	2.0
Aliss (Inra Ivry) (19)(-28)	58	40.8	1.9
Autre (Grenoble 2-Inra) (38)(-21)	59	37.8	1.8
Centre (Pau) (17)(-2)	60	37.1	1.8
Mona-Tsv *(Inra Ivry)* (19)(+38)	61	35.0	1.7
Non-Crem (Caen-Rennes 1) (39)(-7)	62	34.8	1.7
Lef *(Inra Nancy)* (7)(+23)	63	34.4	1.6
Centre *(Paris 12)* (30)(+6)	64	33.2	1.6
Autre (Montpellier 1-Inra) (21)(-9)	65	31.4	1.5
Cesaer (Inra Dijon) (11)(-24)	66	30.4	1.5
Centre *(Paris 7)* (10)(+27)	67	29.5	1.4
Centre (La Reunion) (19)(-17)	68	28.7	1.4
Centre (Mulhouse) (9)(-9)	69	26.6	1.3
Centre (Poitiers) (27)(-5)	70	26.1	1.2
Lest-Eco *(Aix Marseille 2-3)* (13)(+20)	71	25.8	1.2
Centre (Lille 3) (11)(-3)	72	23.5	1.1
Centre (Perpignan) (12)(-38)	73	22.6	1.1
Centre (Angers) (18)(-3)	74	22.3	1.1
Non-Cepn *(Paris 13)* (12)(+5)	75	21.8	1.0
Centre (Ant. Guy.) (22)(-14)	76	20.3	1.0
Centre (Toulon) (11)(+2)	77	19.5	0.9
Centre (Littoral) (13)(-31)	78	19.2	0.9
Centre (Marne La Vallee) (11)(-26)	79	18.8	0.9
Centre *(Valenciennes)* (8)(+13)	80	18.0	0.9
Autre *(Dijon)* (11)(+7)	81	17.8	0.9
Centre (Metz) (13)(-25)	82	17.6	0.8
Centre (Tours) (12)(-15)	83	16.9	0.8
Non-Beta (Nancy 2-Strasb. 1) (24)(-2)	84	16.0	0.8
Moisa (Montpellier 1-Inra) (8)(-10)	85	15.8	0.8
Centre (Brest) (21)(-23)	86	14.9	0.7
Centre (Le Havre) (15)(-4)	87	14.2	0.7
Centre (Amiens) (22)(-7)	88	12.6	0.6
Centre (Limoges) (18)(-15)	89	12.4	0.6
Centre *(Toulouse 2)* (10)(+14)	90	12.4	0.6
Centre *(Toulouse 3)* (10)(+9)	91	12.2	0.6
Centre (La Rochelle) (5)(-20)	92	11.6	0.6
Centre (Rennes 2) (7)(-1)	93	11.4	0.5
Centre *(Lyon 1)* (5)(+4)	94	10.9	0.5
Non-Gredeg (Nice) (12)(-11)	95	10.7	0.5
Centre (Montpellier 3) (10)(-19)	96	9.7	0.5
Centre (Cnam) (7)(-7)	97	8.9	0.4
Centre (Paris 5) (11)(-26)	98	7.5	0.4
Centre (Artois) (13)(-3)	99	7.2	0.3
Centre (Corte) (9)(-5)	100	4.8	0.2
Non-Economix (Paris 10) (15)(-4)	101	4.0	0.2
Centre (Bretagne Sud) (9)(-18)	102	3.7	0.2
Centre (Lyon 3) (9)(-1)	103	2.8	0.1
Non-Cerdi (Clermont 1) (6)(0)	104	0.1	0

centre	rg.	p.c.	nor.
Non-Gate *(Lyon 2)* (43)(+26)	47	2.42	6.8
Centre *(Valenciennes)* (8)(+24)	48	2.40	6.8
Centre (La Rochelle) (5)(-23)	49	2.32	6.5
Gredeg *(Nice)* (71)(+19)	50	2.29	6.4
Centre (Pau) (17)(-1)	51	2.25	6.3
Leo (Orleans) (30)(-27)	52	2.22	6.3
Aliss (Inra Ivry) (19)(-30)	53	2.21	6.2
Centre *(Lyon 1)* (5)(+17)	54	2.19	6.2
Centre (St Etienne) (22)(-8)	55	2.16	6.1
Centre (Lille 3) (11)(-11)	56	2.14	6.0
Crem (Caen-Rennes 1) (82)(-16)	57	2.07	5.8
Lest-Eco *(Aix Marseille 2-3)* (13)(+32)	58	1.98	5.6
Centre (Perpignan) (12)(-42)	59	1.97	5.5
Moisa (Montpellier 1-Inra) (8)(-15)	60	1.97	5.6
Mona-Tsv *(Inra Ivry)* (19)(+40)	61	1.89	5.3
Centre *(Rouen)* (22)(+26)	62	1.87	5.3
Non-Cepn *(Paris 13)* (12)(+3)	63	1.82	5.1
Centre (Marne La Vallee) (11)(-37)	64	1.79	5.0
Non-Ermes (Paris 2) (24)(-23)	65	1.78	5.0
Non-Gretha (Bordeaux 4) (35)(+2)	66	1.78	5.0
Centre (Toulon) (11)(-6)	67	1.77	5.0
Lepii *(Grenoble 2-Inra)* (29)(+26)	68	1.65	4.7
Centre (Rennes 2) (7)(-1)	69	1.63	4.6
Centre (Paris 8) (27)(-19)	70	1.62	4.6
Centre (Reims) (32)(-5)	71	1.62	4.6
Autre *(Dijon)* (11)(+4)	72	1.61	4.6
Centre (La Reunion) (19)(-27)	73	1.51	4.3
Autre (Montpellier 1-Inra) (21)(-18)	74	1.49	4.2
Autre *(Pse-Paris 1)* (33)(+12)	75	1.49	4.2
Centre (Littoral) (13)(-43)	76	1.48	4.2
Centre (Metz) (13)(-38)	77	1.41	4.0
Centre (Tours) (12)(-28)	78	1.41	4.0
Autre (Lille 1-Poly. Lille) (31)(-36)	79	1.37	3.9
Centre *(Toulouse 2)* (10)(+22)	80	1.30	3.7
Centre (Cnam) (7)(-21)	81	1.28	3.6
Lem (Lille 1-Poly. Lille) (87)(0)	82	1.27	3.6
Centre (Angers) (18)(-8)	83	1.24	3.5
Centre *(Toulouse 3)* (10)(+15)	84	1.22	3.4
Centre *(Paris 12)* (30)(+4)	85	1.13	3.2
Autre *(Aix Marseille 2-3)* (57)(+10)	86	1.07	3.0
Centre (Montpellier 3) (10)(-35)	87	1.02	2.9
Autre (Grenoble 2-Inra) (38)(-25)	88	1.01	2.8
Centre (Poitiers) (27)(-10)	89	0.97	2.7
Centre (Ant. Guy.) (22)(-26)	90	0.94	2.7
Centre (Le Havre) (15)(-9)	91	0.94	2.7
Non-Crem (Caen-Rennes 1) (39)(-12)	92	0.89	2.5
Non-Gredeg (Nice) (12)(-19)	93	0.89	2.5
Centre (Brest) (21)(-32)	94	0.73	2.1
Centre (Paris 5) (11)(-42)	95	0.72	2.0
Centre (Limoges) (18)(-18)	96	0.69	2.0
Non-Beta (Nancy 2-Strasb. 1) (24)(-4)	97	0.68	1.9
Centre (Amiens) (22)(-7)	98	0.57	1.6
Centre (Artois) (13)(-3)	99	0.57	1.6
Centre (Corte) (9)(-13)	100	0.53	1.5
Centre (Bretagne Sud) (9)(-42)	101	0.41	1.2
Centre (Lyon 3) (9)(-2)	102	0.31	0.9
Non-Economix (Paris 10) (15)(-5)	103	0.28	0.8
Non-Cerdi (Clermont 1) (6)(0)	104	0.01	0

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.A.2. Universités, T=Carrière

TABLE 1.25 – Universités, Citations totales, T=Carrière

université	rg.	tot.	nor.	université	rg.	p.c.	nor.
Tse-Toulouse 1 (125)(+1)	1	2556.0	100.0	**Iep Paris** (9)(0)	1	33.62	100.0
Pse-Paris 1 (214)(-1)	2	2294.5	89.8	**Tse-Toulouse 1** (125)(0)	2	20.51	61.0
Crest-Ensae (67)(0)	3	822.5	32.2	**Crest-Ensae** (67)(0)	3	12.26	36.5
Hec (75)(+1)	4	756.3	29.6	**Inra Vers-Grig** (12)(+2)	4	11.65	34.7
Aix Marseille 2-3 (114)(-1)	5	530.7	20.8	**Pse-Paris 1** (214)(-1)	5	10.73	31.9
Paris 9 (124)(+5)	6	370.5	14.5	***Hec*** (75)(+6)	6	10.10	30.0
Ec. Polytechnique (33)(0)	7	337.0	13.2	**Ec. Polytechnique** (33)(-2)	7	10.09	30.0
Paris 10 (80)(-2)	8	323.3	12.7	**Ens Cachan** (7)(-1)	8	7.45	22.2
Iep Paris (9)(+9)	9	285.7	11.2	**Inra Rennes** (12)(-1)	9	7.33	21.8
Nancy 2-Strasbourg 1 (95)(-2)	10	285.2	11.2	***Cired*** (14)(+15)	10	7.02	20.9
Grenoble 2-Inra (128)(+4)	11	276.6	10.8	**Cergy Pontoise** (37)(-1)	11	6.08	18.1
Lille 1-Polytech Lille (153)(0)	12	241.0	9.4	***Inra Nancy*** (7)(+25)	12	5.29	15.7
Lyon 2 (70)(+1)	13	233.7	9.1	**Besancon** (24)(+2)	13	5.13	15.3
Cergy Pontoise (37)(-4)	14	225.1	8.8	*Strasbourg 3* (13)(-3)	14	4.98	14.8
Caen-Rennes 1 (121)(-6)	15	203.4	8.0	*Le Mans* (18)(-6)	15	4.86	14.5
Dijon (65)(+15)	16	185.4	7.3	***Lille 2*** (13)(+11)	16	4.65	13.8
Nice (83)(+6)	17	171.9	6.7	*Aix Marseille 2-3* (114)(-3)	17	4.64	13.8
Bordeaux 4 (72)(-5)	18	163.3	6.4	***Clermont 1*** (32)(+6)	18	4.11	12.2
Montpellier 1-Inra (62)(-3)	19	144.1	5.6	**Paris 10** (80)(-1)	19	4.04	12.0
Inra Vers-Grig (12)(+4)	20	139.8	5.5	***Versailles St Quentin*** (24)(+26)	20	3.83	11.4
Clermont 1 (32)(-1)	21	131.6	5.2	*Nantes* (23)(-4)	21	3.62	10.8
Paris 13 (45)(+8)	22	130.8	5.1	***Paris 11*** (36)(+45)	22	3.33	9.9
Besancon (24)(-1)	23	123.2	4.8	***Lyon 2*** (70)(+11)	23	3.32	9.9
Paris 11 (36)(+25)	24	120.0	4.7	***Chambery*** (15)(+21)	24	3.17	9.4
Paris 2 (41)(-6)	25	113.1	4.4	**Mulhouse** (9)(-2)	25	3.12	9.3
Cired (14)(+10)	26	94.7	3.7	*Evry* (18)(-7)	26	3.06	9.1
Versailles St Quentin (24)(+15)	27	91.9	3.6	**Nancy 2-Strasbourg 1** (95)(+2)	27	3.02	9.0
Le Mans (18)(-7)	28	88.6	3.5	***Paris 9*** (124)(+19)	28	2.98	8.9
Inra Rennes (12)(-3)	29	87.9	3.4	***Paris 7*** (10)(+37)	29	2.95	8.8
Nantes (23)(-5)	30	81.5	3.2	***Paris 13*** (45)(+18)	30	2.94	8.7
Inra Ivry (37)(-4)	31	75.8	3.0	***Dijon*** (65)(+32)	31	2.86	8.5
Orleans (34)(-15)	32	68.5	2.7	*Paris 2* (41)(-4)	32	2.79	8.3
Strasbourg 3 (13)(-4)	33	63.7	2.5	*Inra Dijon* (11)(-17)	33	2.76	8.2
Lille 2 (13)(+6)	34	60.4	2.4	***Valenciennes*** (8)(+25)	34	2.40	7.1
Evry (18)(-7)	35	55.0	2.2	*La Rochelle* (5)(-15)	35	2.32	6.9
Ens Cachan (7)(-3)	36	54.6	2.1	*Montpellier 1-Inra* (62)(-4)	36	2.32	6.9
Reims (32)(+2)	37	51.1	2.0	*Bordeaux 4* (72)(-5)	37	2.28	6.8
Chambery (15)(+13)	38	47.6	1.9	***Pau*** (17)(+3)	38	2.25	6.7
St Etienne (22)(-2)	39	47.5	1.9	***Lyon 1*** (5)(+19)	39	2.19	6.5
Paris 8 (27)(-6)	40	43.7	1.7	***Grenoble 2-Inra*** (128)(+19)	40	2.16	6.4
Rouen (22)(+19)	41	41.1	1.6	*St Etienne* (22)(-3)	41	2.16	6.4
Pau (17)(+3)	42	37.1	1.5	*Lille 3* (11)(-9)	42	2.14	6.4
Inra Nancy (7)(+22)	43	34.4	1.3	***Nice*** (83)(+14)	43	2.08	6.2
Paris 12 (30)(+10)	44	33.2	1.3	*Inra Ivry* (37)(-8)	44	2.05	6.1
Inra Dijon (11)(-10)	45	30.4	1.2	*Orleans* (34)(-23)	45	2.02	6.0
Paris 7 (10)(+23)	46	29.5	1.2	*Perpignan* (12)(-33)	46	1.97	5.9
La Reunion (19)(-6)	47	28.7	1.1	***Rouen*** (22)(+22)	47	1.87	5.6
Mulhouse (9)(-2)	48	26.6	1.0	*Marne La Vallee* (11)(-27)	48	1.79	5.3
Poitiers (27)(+1)	49	26.1	1.0	**Toulon** (11)(+2)	49	1.77	5.3
Lille 3 (11)(+3)	50	23.5	0.9	*Caen-Rennes 1* (121)(-11)	50	1.69	5.0
Perpignan (12)(-19)	51	22.6	0.9	***Rennes 2*** (7)(+5)	51	1.63	4.8
Angers (18)(+3)	52	22.3	0.9	*Paris 8* (27)(-10)	52	1.62	4.8
Antilles Guyane (22)(-6)	53	20.3	0.8	**Reims** (32)(+2)	53	1.62	4.8
Toulon (11)(+7)	54	19.5	0.8	**Lille 1-Polytech Lille** (153)(+1)	54	1.57	4.7
Littoral (13)(-17)	55	19.2	0.8	*La Reunion* (19)(-21)	55	1.51	4.5
Marne La Vallee (11)(-13)	56	18.8	0.7	*Littoral* (13)(-30)	56	1.48	4.4
Valenciennes (8)(+11)	57	18.0	0.7	*Metz* (13)(-27)	57	1.41	4.2
Metz (13)(-14)	58	17.6	0.7	*Tours* (12)(-18)	58	1.41	4.2
Tours (12)(-7)	59	16.9	0.7	***Toulouse 2*** (10)(+16)	59	1.30	3.9
Brest (21)(-12)	60	14.9	0.6	*Cnam* (7)(-10)	60	1.28	3.8

suite page suivante

suite de la page précédente

université	rg.	tot.	nor.
Le Havre (15)(+2)	61	14.2	0.6
Amiens (22)(0)	62	12.6	0.5
Limoges (18)(-5)	63	12.4	0.5
Toulouse 2 (10)(+12)	64	12.4	0.5
Toulouse 3 (10)(+8)	65	12.2	0.5
La Rochelle (5)(-10)	66	11.6	0.5
Rennes 2 (7)(0)	67	11.4	0.5
Lyon 1 (5)(+4)	68	10.9	0.4
Montpellier 3 (10)(-10)	69	9.7	0.4
Cnam (7)(-4)	70	8.9	0.4
Paris 5 (11)(-15)	71	7.5	0.3
Artois (13)(-1)	72	7.2	0.3
Corte (9)(-3)	73	4.8	0.2
Bretagne Sud (9)(-10)	74	3.7	0.2
Lyon 3 (9)(-1)	75	2.8	0.1

université	rg.	p.c.	nor.
Angers (18)(0)	61	1.24	3.7
Toulouse 3 (10)(+11)	62	1.22	3.6
Paris 12 (30)(+7)	63	1.13	3.4
Montpellier 3 (10)(-21)	64	1.02	3.0
Poitiers (27)(-2)	65	0.97	2.9
Antilles Guyane (22)(-13)	66	0.94	2.8
Le Havre (15)(-1)	67	0.94	2.8
Brest (21)(-16)	68	0.73	2.2
Paris 5 (11)(-25)	69	0.72	2.1
Limoges (18)(-8)	70	0.69	2.1
Amiens (22)(0)	71	0.57	1.7
Artois (13)(+1)	72	0.57	1.7
Corte (9)(-6)	73	0.53	1.6
Bretagne Sud (9)(-25)	74	0.41	1.2
Lyon 3 (9)(-1)	75	0.31	0.9

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.A.3. Centres, moins de 50 ans

TABLE 1.26 – Centres, Moins de 50 ans, Citations totales, T=Dégressif

centre	**rg.**	**tot.**	**nor.**
Pjse (Pse-Paris 1) (23)(+2)	1	3079.8	100.0
Gremaq (Tse-Toulouse 1) (41)(-1)	2	2561.7	83.2
Greghec (Hec) (61)(+1)	3	2188.3	71.1
Ces (Pse-Paris 1) (80)(-2)	4	2031.4	66.0
Lerna (Tse-Toulouse 1) (14)(+1)	5	1186.0	38.5
Greqam (Aix Marseille 2-3) (27)(-1)	6	1119.0	36.3
Economix (Paris 10) (44)(+2)	7	839.0	27.2
Non-Grecsta (Crest-Ensae) (15)(+2)	8	801.1	26.0
Beta (Nancy 2-Strasb. 1) (52)(-1)	9	772.2	25.1
Thema (Cergy Pontoise) (32)(-3)	10	732.7	23.8
Preg (Ec. Polytechnique) (19)(0)	11	632.1	20.5
Grecsta (Crest-Ensae) (15)(+1)	12	500.3	16.3
Crem (Caen-Rennes 1) (64)(-1)	13	461.6	15.0
Eco. Pub. *(Inra Vers-Grig)* (8)(+5)	14	446.2	14.5
Gate *(Lyon 2)* (19)(+5)	15	391.6	12.7
Cerdi (Clermont 1) (18)(+2)	16	361.8	11.8
Drm *(Paris 9)* (43)(+8)	17	350.9	11.4
Lem (Lille 1-Poly. Lille) (70)(-3)	18	339.5	11.0
Smart *(Inra Rennes)* (11)(+3)	19	305.2	9.9
Centre *(Besancon)* (16)(+6)	20	287.3	9.3
Clerse-Eco *(Lille 1-Poly. Lille)* (23)(+10)	21	282.4	9.2
Centre *(Cired)* (7)(+20)	22	243.6	7.9
Non-Drm *(Paris 9)* (27)(+5)	23	242.8	7.9
Centre *(Lille 2)* (10)(+22)	24	231.6	7.5
Ermes *(Paris 2)* (10)(+7)	25	222.7	7.2
Leo (Orleans) (19)(+1)	26	210.4	6.8
Leg *(Dijon)* (32)(+25)	27	205.2	6.7
Centre *(Chambery)* (12)(+26)	28	204.4	6.6
Gretha (Bordeaux 4) (26)(-5)	29	191.4	6.2
Centre (Nantes) (17)(-13)	30	186.1	6.0
Centre (Strasbourg 3) (11)(-2)	31	182.8	5.9
Centre (Le Mans) (11)(-18)	32	181.2	5.9
Gredeg (Nice) (34)(+1)	33	179.1	5.8
Centre *(Vers. St Quentin)* (17)(+16)	34	176.6	5.7
Non-Gretha *(Bordeaux 4)* (17)(+21)	35	171.5	5.6
Cerag *(Grenoble 2-Inra)* (35)(+8)	36	169.0	5.5
Non-Gate *(Lyon 2)* (17)(+29)	37	168.6	5.5
Centre (Evry) (16)(-8)	38	167.1	5.4
Centre *(Pau)* (9)(+23)	39	156.1	5.1
Centre (Reims) (19)(0)	40	150.6	4.9
Autre (Tse-Toulouse 1) (7)(-8)	41	143.3	4.7
Gael (Grenoble 2-Inra) (10)(-4)	42	143.0	4.6
Non-Cermes *(Paris 11)* (8)(+10)	43	141.9	4.6
Lirhe-Eco *(Tse-Toulouse 1)* (12)(+42)	44	141.6	4.6
Aliss (Inra Ivry) (13)(-22)	45	141.5	4.6
Centre (St Etienne) (17)(-10)	46	141.0	4.6
Lepii *(Grenoble 2-Inra)* (17)(+28)	47	133.2	4.3
Centre (Paris 8) (13)(-11)	48	127.4	4.1
Lameta (Montpellier 1-Inra) (17)(-33)	49	126.8	4.1
Autre (Lille 1-Poly. Lille) (18)(-29)	50	125.9	4.1
Cepn (Paris 13) (18)(-6)	51	122.4	4.0
Centre *(Rouen)* (12)(+25)	52	97.5	3.2
Lest-Eco *(Aix Marseille 2-3)* (9)(+35)	53	91.1	3.0
Non-Ermes (Paris 2) (10)(-6)	54	87.7	2.9
Cesaer (Inra Dijon) (8)(-14)	55	81.0	2.6
Autre (Montpellier 1-Inra) (6)(-13)	56	76.2	2.5
Centre *(Paris 12)* (7)(+28)	57	75.9	2.5
Centre (Perpignan) (8)(-23)	58	74.8	2.4
Non-Crem (Caen-Rennes 1) (25)(-12)	59	72.3	2.4
Mona-Tsv *(Inra Ivry)* (7)(+30)	60	70.7	2.3

centre	**rg.**	**p.c.**	**nor.**
Pjse (Pse-Paris 1) (23)(0)	1	133.61	100.0
Lerna (Tse-Toulouse 1) (14)(+1)	2	84.72	63.4
Gremaq (Tse-Toulouse 1) (41)(-1)	3	62.03	46.4
Eco. Pub. (Inra Vers-Grig) (8)(+2)	4	55.78	41.8
Non-Grecsta (Crest-Ensae) (15)(0)	5	54.87	41.1
Greqam (Aix Marseille 2-3) (27)(-2)	6	42.22	31.6
Centre *(Cired)* (7)(+17)	7	37.48	28.1
Greghec *(Hec)* (61)(+5)	8	35.87	26.9
Preg (Ec. Polytechnique) (19)(-2)	9	33.89	25.4
Grecsta (Crest-Ensae) (15)(-1)	10	32.92	24.6
Smart *(Inra Rennes)* (11)(+3)	11	27.74	20.8
Ces (Pse-Paris 1) (80)(-1)	12	25.50	19.1
Ermes *(Paris 2)* (10)(+4)	13	23.44	17.6
Thema (Cergy Pontoise) (32)(-4)	14	23.26	17.4
Centre *(Lille 2)* (10)(+25)	15	23.16	17.3
Autre (Tse-Toulouse 1) (7)(-4)	16	20.47	15.3
Gate *(Lyon 2)* (19)(+13)	17	20.19	15.1
Cerdi *(Clermont 1)* (18)(+8)	18	20.10	15.0
Economix *(Paris 10)* (44)(+3)	19	19.27	14.4
Non-Cermes *(Paris 11)* (8)(+15)	20	18.92	14.2
Centre *(Besancon)* (16)(+10)	21	17.96	13.4
Centre *(Pau)* (9)(+24)	22	17.34	13.0
Centre *(Chambery)* (12)(+34)	23	17.04	12.8
Centre (Strasbourg 3) (11)(-9)	24	16.77	12.6
Centre (Le Mans) (11)(-17)	25	16.10	12.1
Beta (Nancy 2-Strasb. 1) (52)(+2)	26	14.85	11.1
Gael *(Grenoble 2-Inra)* (10)(+7)	27	14.30	10.7
Centre (La Rochelle) (5)(-7)	28	13.98	10.5
Autre (Montpellier 1-Inra) (6)(-10)	29	12.70	9.5
Lef *(Inra Nancy)* (6)(+31)	30	12.44	9.3
Lirhe-Eco *(Tse-Toulouse 1)* (12)(+56)	31	12.32	9.2
Clerse-Eco *(Lille 1-Poly. Lille)* (23)(+17)	32	12.28	9.2
Centre *(Paris 12)* (7)(+45)	33	11.68	8.7
Aliss (Inra Ivry) (13)(-16)	34	11.32	8.5
Centre (Nantes) (17)(-15)	35	11.28	8.4
Leo (Orleans) (19)(+1)	36	11.08	8.3
Centre (Montpellier 3) (6)(-12)	37	10.89	8.2
Mona-Tsv *(Inra Ivry)* (7)(+52)	38	10.88	8.1
Centre (Evry) (16)(-6)	39	10.78	8.1
Centre *(Vers. St Quentin)* (17)(+26)	40	10.39	7.8
Non-Gate *(Lyon 2)* (17)(+36)	41	10.22	7.7
Centre (Paris 8) (13)(-4)	42	10.19	7.6
Cesaer (Inra Dijon) (8)(-14)	43	10.12	7.6
Lest-Eco *(Aix Marseille 2-3)* (9)(+45)	44	10.12	7.6
Non-Gretha *(Bordeaux 4)* (17)(+23)	45	10.09	7.6
Cermes *(Paris 11)* (6)(+43)	46	9.79	7.3
Centre (Mulhouse) (6)(-20)	47	9.42	7.1
Centre (Perpignan) (8)(-32)	48	9.35	7.0
Non-Ermes (Paris 2) (10)(-10)	49	9.23	6.9
Non-Drm *(Paris 9)* (27)(+4)	50	8.99	6.7
Centre *(Rennes 2)* (6)(+13)	51	8.94	6.7
Centre (Tours) (6)(-16)	52	8.37	6.3
Centre (St Etienne) (17)(-2)	53	8.30	6.2
Drm *(Paris 9)* (43)(+15)	54	8.16	6.1
Centre *(Rouen)* (12)(+24)	55	8.12	6.1
Centre *(Reims)* (19)(+4)	56	7.92	5.9
Lepii *(Grenoble 2-Inra)* (17)(+28)	57	7.83	5.9
Centre (Lille 3) (6)(-5)	58	7.70	5.8
Gretha (Bordeaux 4) (26)(-11)	59	7.51	5.6
Lameta (Montpellier 1-Inra) (17)(-37)	60	7.46	5.6

suite page suivante

suite de la page précédente

centre	rg.	tot.	nor.
Centre (La Rochelle) (5)(-6)	61	69.9	2.3
Lef *(Inra Nancy)* (6)(+20)	62	68.4	2.2
Centre *(Angers)* (14)(+6)	63	66.1	2.2
Centre (Montpellier 3) (6)(-7)	64	59.9	2.0
Centre (Limoges) (9)(0)	65	57.5	1.9
Autre *(Pse-Paris 1)* (13)(+4)	66	56.9	1.9
Centre (La Reunion) (15)(-16)	67	56.8	1.9
Cermes *(Paris 11)* (6)(+21)	68	53.9	1.8
Centre *(Rennes 2)* (6)(+11)	69	53.6	1.7
Centre (Littoral) (12)(-21)	70	53.5	1.7
Centre (Mulhouse) (6)(-10)	71	51.8	1.7
Centre (Marne La Vallee) (10)(-14)	72	51.5	1.7
Autre (Aix Marseille 2-3) (25)(-9)	73	50.3	1.6
Centre (Lille 3) (6)(+2)	74	46.2	1.5
Centre (Poitiers) (17)(-11)	75	46.2	1.5
Centre (Tours) (6)(-9)	76	46.0	1.5
Centre (Ant. Guy.) (13)(-18)	77	41.4	1.3
Centre (Le Havre) (11)(+1)	78	40.2	1.3
Centre (Metz) (7)(-5)	79	37.6	1.2
Non-Cepn (Paris 13) (7)(-9)	80	36.9	1.2
Autre (Grenoble 2-Inra) (13)(-42)	81	36.8	1.2
Centre (Toulon) (5)(-9)	82	35.9	1.2
Centre (Brest) (12)(-23)	83	30.6	1.0
Centre (Paris 5) (6)(-12)	84	28.2	0.9
Non-Beta (Nancy 2-Strasb. 1) (7)(-17)	85	28.1	0.9
Centre (Lyon 1) (5)(-2)	86	24.5	0.8
Centre (Artois) (7)(-9)	87	17.9	0.6
Centre (Amiens) (11)(-7)	88	16.8	0.6
Centre (Corte) (6)(-2)	89	14.0	0.5
Centre (Bretagne Sud) (7)(-7)	90	10.3	0.3

centre	rg.	p.c.	nor.
Crem (Caen-Rennes 1) (64)(-16)	61	7.27	5.4
Centre (Toulon) (5)(-20)	62	7.17	5.4
Autre (Lille 1-Poly. Lille) (18)(-31)	63	6.99	5.2
Cepn (Paris 13) (18)(0)	64	6.99	5.2
Centre (Limoges) (9)(-13)	65	6.77	5.1
Leg *(Dijon)* (32)(+18)	66	6.51	4.9
Centre (Metz) (7)(-11)	67	5.79	4.3
Centre (Marne La Vallee) (10)(-21)	68	5.42	4.1
Gredeg (Nice) (34)(+2)	69	5.27	3.9
Non-Cepn (Paris 13) (7)(-14)	70	5.27	4.0
Centre (Paris 5) (6)(-28)	71	5.12	3.8
Centre (Lyon 1) (5)(-5)	72	4.91	3.7
Cerag *(Grenoble 2-Inra)* (35)(+9)	73	4.90	3.7
Lem (Lille 1-Poly. Lille) (70)(+1)	74	4.85	3.6
Centre (Angers) (14)(-2)	75	4.72	3.5
Autre (Pse-Paris 1) (13)(-4)	76	4.55	3.4
Centre (Littoral) (12)(-27)	77	4.45	3.3
Non-Beta (Nancy 2-Strasb. 1) (7)(-34)	78	4.33	3.2
Centre (La Reunion) (15)(-17)	79	3.79	2.8
Centre (Le Havre) (11)(0)	80	3.66	2.7
Centre (Ant. Guy.) (13)(-22)	81	3.31	2.5
Autre (Grenoble 2-Inra) (13)(-41)	82	2.94	2.2
Non-Crem (Caen-Rennes 1) (25)(-7)	83	2.89	2.2
Centre (Poitiers) (17)(-10)	84	2.72	2.0
Centre (Brest) (12)(-27)	85	2.66	2.0
Centre (Artois) (7)(-21)	86	2.56	1.9
Centre (Corte) (6)(-4)	87	2.33	1.7
Autre (Aix Marseille 2-3) (25)(-2)	88	2.01	1.5
Centre (Amiens) (11)(-8)	89	1.53	1.1
Centre (Bretagne Sud) (7)(-20)	90	1.48	1.1

La colonne "rg." donne le rang, la colonne "tot." donne le score total, "nor." le score normalisé par rapport à celui du premier classé, "p.c." donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

TABLE 1.27 – Centres, Moins de 50 ans, Indice G, T=Dégressif

centre	rg.	tot.	nor.
Pjse (Pse-Paris 1) (23)(+2)	1	77.6	100.0
Gremaq (Tse-Toulouse 1) (41)(-1)	2	69.1	89.0
Greghec (Hec) (61)(+1)	3	66.6	85.8
Ces (Pse-Paris 1) (80)(-2)	4	59.6	76.9
Non-Grecsta *(Crest-Ensae)* (15)(+5)	5	53.7	69.3
Greqam (Aix Marseille 2-3) (27)(-1)	6	45.3	58.4
Grecsta *(Crest-Ensae)* (15)(+6)	7	41.7	53.7
Lerna (Tse-Toulouse 1) (14)(-1)	8	41.7	53.8
Economix (Paris 10) (44)(0)	9	41.1	53.0
Beta (Nancy 2-Strasb. 1) (52)(-2)	10	39.6	51.0
Preg (Ec. Polytechnique) (19)(0)	11	39.4	50.7
Eco. Pub. *(Inra Vers-Grig)* (8)(+7)	12	32.6	42.0
Thema (Cergy Pontoise) (32)(-6)	13	31.9	41.2
Gate *(Lyon 2)* (19)(+6)	14	31.4	40.5
Centre (Le Mans) (11)(-1)	15	30.4	39.2
Crem (Caen-Rennes 1) (64)(-4)	16	28.9	37.2
Ermes *(Paris 2)* (10)(+15)	17	28.3	36.5
Drm *(Paris 9)* (43)(+7)	18	25.3	32.7
Smart *(Inra Rennes)* (11)(+3)	19	25.1	32.3
Centre *(Cired)* (7)(+22)	20	24.2	31.3
Cerdi (Clermont 1) (18)(-3)	21	23.8	30.7
Lem (Lille 1-Poly. Lille) (70)(-7)	22	23.7	30.6

centre	rg.	p.c.	nor.
Pjse (Pse-Paris 1) (23)(0)	1	9.44	100.0
Lerna (Tse-Toulouse 1) (14)(+1)	2	7.48	79.3
Centre *(Cired)* (7)(+21)	3	6.24	66.2
Eco. Pub. (Inra Vers-Grig) (8)(+2)	4	6.08	64.4
Gremaq (Tse-Toulouse 1) (41)(-3)	5	5.74	60.8
Non-Grecsta (Crest-Ensae) (15)(-1)	6	5.30	56.1
Greqam (Aix Marseille 2-3) (27)(-3)	7	5.14	54.5
Smart *(Inra Rennes)* (11)(+6)	8	4.61	48.9
Preg (Ec. Polytechnique) (19)(-2)	9	4.19	44.4
Centre *(Chambery)* (12)(+47)	10	3.83	40.6
Grecsta (Crest-Ensae) (15)(-2)	11	3.79	40.1
Ces (Pse-Paris 1) (80)(-1)	12	3.66	38.8
Thema (Cergy Pontoise) (32)(-3)	13	3.63	38.5
Economix *(Paris 10)* (44)(+8)	14	3.55	37.6
Non-Cermes *(Paris 11)* (8)(+20)	15	3.51	37.2
Centre (Le Mans) (11)(-8)	16	3.40	36.0
Greghec (Hec) (61)(-4)	17	3.31	35.1
Cerdi *(Clermont 1)* (18)(+8)	18	3.29	34.8
Autre (Tse-Toulouse 1) (7)(-7)	19	3.26	34.5
Autre (Montpellier 1-Inra) (6)(-1)	20	3.23	34.2
Centre *(Montpellier 3)* (6)(+4)	21	3.18	33.7
Lef *(Inra Nancy)* (6)(+39)	22	3.17	33.5

suite page suivante

suite de la page précédente

centre	**rg.**	**tot.**	**nor.**
Centre *(Lille 2)* (10)(+23)	23	22.9	29.5
Centre *(Chambery)* (12)(+30)	24	22.8	29.4
Centre (Besancon) (16)(+1)	25	22.5	29.0
Lirhe-Eco *(Tse-Toulouse 1)* (12)(+60)	26	22.2	28.6
Non-Gretha *(Bordeaux 4)* (17)(+29)	27	19.9	25.6
Centre (Nantes) (17)(-11)	28	19.7	25.4
Gael *(Grenoble 2-Inra)* (10)(+9)	29	19.5	25.2
Clerse-Eco (Lille 1-Poly. Lille) (23)(+1)	30	19.4	25.1
Gredeg *(Nice)* (34)(+4)	31	19.4	25.0
Centre *(Vers. St Quentin)* (17)(+18)	32	19.0	24.5
Leg *(Dijon)* (32)(+19)	33	18.6	24.0
Non-Cermes *(Paris 11)* (8)(+20)	34	18.6	24.0
Centre (Evry) (16)(-5)	35	18.4	23.8
Centre *(Reims)* (19)(+4)	36	18.3	23.6
Non-Drm (Paris 9) (27)(-9)	37	18.0	23.2
Centre *(Pau)* (9)(+24)	38	17.7	22.8
Leo (Orleans) (19)(-12)	39	17.6	22.7
Aliss (Inra Ivry) (13)(-17)	40	17.1	22.1
Gretha (Bordeaux 4) (26)(-17)	41	17.0	21.9
Centre (Strasbourg 3) (11)(-13)	42	16.7	21.5
Autre (Tse-Toulouse 1) (7)(-10)	43	15.6	20.1
Centre (St Etienne) (17)(-7)	44	15.6	20.1
Centre *(Angers)* (14)(+24)	45	15.3	19.8
Lameta (Montpellier 1-Inra) (17)(-29)	46	15.3	19.7
Non-Gate *(Lyon 2)* (17)(+19)	47	14.6	18.9
Cerag (Grenoble 2-Inra) (35)(-4)	48	14.3	18.5
Autre (Lille 1-Poly. Lille) (18)(-28)	49	13.8	17.8
Non-Ermes (Paris 2) (10)(-1)	50	13.8	17.8
Autre (Montpellier 1-Inra) (6)(-8)	51	13.6	17.5
Lest-Eco *(Aix Marseille 2-3)* (9)(+36)	52	13.3	17.2
Cepn (Paris 13) (18)(-8)	53	13.2	17.1
Centre *(Rouen)* (12)(+23)	54	13.1	16.8
Non-Crem (Caen-Rennes 1) (25)(-8)	55	12.7	16.4
Centre (Paris 8) (13)(-19)	56	12.4	16.0
Lef *(Inra Nancy)* (6)(+26)	57	12.4	16.0
Cesaer (Inra Dijon) (8)(-17)	58	12.3	15.9
Centre *(Paris 12)* (7)(+26)	59	12.1	15.6
Mona-Tsv *(Inra Ivry)* (7)(+31)	60	12.1	15.7
Centre (Perpignan) (8)(-26)	61	12.0	15.4
Lepii *(Grenoble 2-Inra)* (17)(+14)	62	12.0	15.5
Centre (Limoges) (9)(+2)	63	11.8	15.3
Centre (Mulhouse) (6)(-3)	64	11.5	14.9
Centre (Poitiers) (17)(-2)	65	11.2	14.5
Centre (Tours) (6)(+1)	66	10.9	14.0
Autre *(Pse-Paris 1)* (13)(+3)	67	10.5	13.6
Centre (Montpellier 3) (6)(-11)	68	10.4	13.4
Autre (Aix Marseille 2-3) (25)(-5)	69	10.2	13.2
Centre (La Reunion) (15)(-18)	70	10.2	13.2
Centre (Marne La Vallee) (10)(-11)	71	10.2	13.2
Centre *(Le Havre)* (11)(+7)	72	9.9	12.7
Centre (Toulon) (5)(0)	73	9.8	12.6
Cermes *(Paris 11)* (6)(+16)	74	9.8	12.6
Centre (Metz) (7)(-1)	75	9.5	12.2
Centre (Lille 3) (6)(0)	76	9.4	12.2
Centre (La Rochelle) (5)(-22)	77	9.3	12.0
Centre (Rennes 2) (6)(+2)	78	9.2	11.9
Non-Beta (Nancy 2-Strasb. 1) (7)(-11)	79	9.1	11.7
Centre (Ant. Guy.) (13)(-21)	80	8.2	10.6
Centre (Brest) (12)(-20)	81	8.2	10.6
Non-Cepn (Paris 13) (7)(-11)	82	7.7	9.9
Centre (Littoral) (12)(-34)	83	7.6	9.7
Centre (Paris 5) (6)(-12)	84	7.5	9.7
Autre (Grenoble 2-Inra) (13)(-46)	85	7.1	9.1
Centre (Artois) (7)(-8)	86	5.6	7.2

centre	**rg.**	**p.c.**	**nor.**
Gate *(Lyon 2)* (19)(+7)	23	3.15	33.4
Centre (Perpignan) (8)(-8)	24	2.95	31.3
Centre (Strasbourg 3) (11)(-9)	25	2.95	31.3
Centre *(Pau)* (9)(+20)	26	2.93	31.1
Beta (Nancy 2-Strasb. 1) (52)(+1)	27	2.86	30.4
Centre *(Paris 12)* (7)(+50)	28	2.81	29.8
Gael *(Grenoble 2-Inra)* (10)(+5)	29	2.80	29.7
Ermes (Paris 2) (10)(-13)	30	2.69	28.6
Lest-Eco *(Aix Marseille 2-3)* (9)(+57)	31	2.68	28.5
Lameta (Montpellier 1-Inra) (17)(-9)	32	2.58	27.3
Centre *(Rennes 2)* (6)(+31)	33	2.57	27.2
Centre (Evry) (16)(-1)	34	2.50	26.5
Centre (La Rochelle) (5)(-14)	35	2.48	26.3
Clerse-Eco *(Lille 1-Poly. Lille)* (23)(+13)	36	2.43	25.8
Leo (Orleans) (19)(0)	37	2.38	25.2
Lepii *(Grenoble 2-Inra)* (17)(+47)	38	2.34	24.8
Cermes *(Paris 11)* (6)(+50)	39	2.31	24.5
Aliss (Inra Ivry) (13)(-22)	40	2.29	24.2
Lirhe-Eco *(Tse-Toulouse 1)* (12)(+47)	41	2.29	24.3
Centre *(Rouen)* (12)(+37)	42	2.26	24.0
Mona-Tsv *(Inra Ivry)* (7)(+47)	43	2.25	23.8
Non-Cepn *(Paris 13)* (7)(+11)	44	2.21	23.4
Centre (Besancon) (16)(-14)	45	2.19	23.3
Centre (Mulhouse) (6)(-19)	46	2.18	23.1
Centre (Paris 8) (13)(-8)	47	2.18	23.1
Cepn *(Paris 13)* (18)(+15)	48	2.17	23.0
Gretha (Bordeaux 4) (26)(-1)	49	2.14	22.7
Centre (Nantes) (17)(-30)	50	2.13	22.6
Crem (Caen-Rennes 1) (64)(-6)	51	2.12	22.5
Non-Drm (Paris 9) (27)(+2)	52	2.10	22.2
Centre (Lille 2) (10)(-13)	53	2.09	22.2
Cesaer (Inra Dijon) (8)(-25)	54	2.04	21.6
Centre (St Etienne) (17)(-4)	55	2.01	21.4
Centre (Toulon) (5)(-14)	56	2.00	21.2
Centre *(Vers. St Quentin)* (17)(+9)	57	1.92	20.3
Autre (Lille 1-Poly. Lille) (18)(-26)	58	1.90	20.1
Centre (Reims) (19)(+1)	59	1.85	19.6
Non-Ermes (Paris 2) (10)(-20)	60	1.85	19.6
Drm *(Paris 9)* (43)(+8)	61	1.83	19.4
Centre (Tours) (6)(-26)	62	1.79	19.0
Centre (Lille 3) (6)(-10)	63	1.72	18.2
Non-Gretha *(Bordeaux 4)* (17)(+4)	64	1.70	18.0
Non-Gate *(Lyon 2)* (17)(+12)	65	1.69	17.9
Centre (Metz) (7)(-10)	66	1.65	17.5
Gredeg *(Nice)* (34)(+4)	67	1.64	17.4
Centre (Marne La Vallee) (10)(-21)	68	1.61	17.0
Centre *(Le Havre)* (11)(+11)	69	1.56	16.6
Cerag *(Grenoble 2-Inra)* (35)(+12)	70	1.52	16.2
Autre (Pse-Paris 1) (13)(+1)	71	1.45	15.3
Centre (Littoral) (12)(-22)	72	1.44	15.3
Lem (Lille 1-Poly. Lille) (70)(+2)	73	1.37	14.5
Leg *(Dijon)* (32)(+10)	74	1.33	14.1
Centre (La Reunion) (15)(-13)	75	1.30	13.8
Centre (Limoges) (9)(-24)	76	1.28	13.5
Centre (Paris 5) (6)(-34)	77	1.26	13.4
Non-Beta (Nancy 2-Strasb. 1) (7)(-34)	78	1.25	13.3
Autre (Grenoble 2-Inra) (13)(-38)	79	1.23	13.1
Centre (Ant. Guy.) (13)(-21)	80	1.05	11.2
Centre (Corte) (6)(+2)	81	1.03	10.9
Centre (Brest) (12)(-24)	82	0.97	10.3
Non-Crem (Caen-Rennes 1) (25)(-7)	83	0.94	10.0
Centre (Artois) (7)(-19)	84	0.93	9.8
Centre (Lyon 1) (5)(-18)	85	0.92	9.7
Centre (Angers) (14)(-13)	86	0.91	9.7

suite page suivante

suite de la page précédente

centre	rg.	tot.	nor.
Centre (Corte) (6)(0)	87	4.9	6.3
Centre (Lyon 1) (5)(-4)	88	4.6	5.9
Centre (Bretagne Sud) (7)(-6)	89	3.9	5.1
Centre (Amiens) (11)(-9)	90	3.1	4.0

centre	rg.	p.c.	nor.
Centre (Bretagne Sud) (7)(-17)	87	0.78	8.2
Centre (Poitiers) (17)(-13)	88	0.78	8.2
Autre (Aix Marseille 2-3) (25)(-3)	89	0.76	8.1
Centre (Amiens) (11)(-9)	90	0.74	7.8

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.A.4. Universités, moins de 50 ans

TABLE 1.28 – Universités, Moins de 50 ans, Citations totales, T=Dégressif

université	rg.	tot.	nor.
Pse-Paris 1 (115)(+1)	1	5168.0	100.0
Tse-Toulouse 1 (74)(-1)	2	4032.6	78.0
Hec (61)(+1)	3	2188.3	42.3
Crest-Ensae (30)(+1)	4	1301.4	25.2
Aix Marseille 2-3 (61)(-2)	5	1260.4	24.4
Paris 10 (46)(+2)	6	861.8	16.7
Nancy 2-Strasbourg 1 (59)(-1)	7	800.3	15.5
Lille 1-Polytech Lille (111)(+1)	8	747.8	14.5
Cergy Pontoise (33)(-2)	9	734.0	14.2
Ec. Polytechnique (19)(+1)	10	632.1	12.2
Paris 9 (70)(+2)	11	593.7	11.5
Lyon 2 (36)(+4)	12	560.3	10.8
Caen-Rennes 1 (89)(-3)	13	533.9	10.3
Grenoble 2-Inra (74)(0)	14	481.9	9.3
Inra Vers-Grig (8)(+5)	15	446.2	8.6
Bordeaux 4 (43)(+1)	16	363.0	7.0
Clermont 1 (19)(+2)	17	361.8	7.0
Paris 2 (19)(+3)	18	310.4	6.0
Inra Rennes (11)(+3)	19	305.2	5.9
Besancon (16)(+6)	20	287.3	5.6
Dijon (38)(+13)	21	272.5	5.3
Cired (7)(+14)	22	243.6	4.7
Montpellier 1-Inra (26)(-11)	23	232.6	4.5
Lille 2 (10)(+13)	24	231.6	4.5
Orleans (21)(-1)	25	213.0	4.1
Inra Ivry (19)(-3)	26	212.2	4.1
Chambery (12)(+15)	27	204.4	4.0
Nice (38)(-3)	28	201.9	3.9
Paris 11 (13)(+10)	29	195.7	3.8
Nantes (17)(-12)	30	186.1	3.6
Strasbourg 3 (11)(-4)	31	182.8	3.5
Le Mans (11)(-17)	32	181.2	3.5
Versailles St Quentin (17)(+7)	33	176.6	3.4
Evry (16)(-6)	34	167.1	3.2
Paris 13 (25)(-6)	35	159.3	3.1
Pau (9)(+13)	36	156.1	3.0
Reims (19)(-4)	37	150.6	2.9
St Etienne (17)(-7)	38	141.0	2.7
Paris 8 (13)(-7)	39	127.4	2.5
Rouen (12)(+18)	40	97.5	1.9
Inra Dijon (8)(-6)	41	81.0	1.6
Paris 12 (7)(+24)	42	75.9	1.5
Perpignan (8)(-13)	43	74.8	1.5

université	rg.	p.c.	nor.
Inra Vers-Grig (8)(+2)	1	55.78	100.0
Tse-Toulouse 1 (74)(-1)	2	54.64	98.0
Pse-Paris 1 (115)(+1)	3	44.86	80.4
Crest-Ensae (30)(-2)	4	43.67	78.3
Cired (7)(+12)	5	37.48	67.2
Hec (61)(+3)	6	35.87	64.3
Ec. Polytechnique (19)(-2)	7	33.89	60.8
Inra Rennes (11)(+2)	8	27.74	49.7
Lille 2 (10)(+21)	9	23.16	41.5
Cergy Pontoise (33)(-3)	10	22.24	39.9
Aix Marseille 2-3 (61)(-3)	11	20.83	37.4
Clermont 1 (19)(+7)	12	19.04	34.1
Paris 10 (46)(+2)	13	18.72	33.6
Besancon (16)(+10)	14	17.96	32.2
Pau (9)(+19)	15	17.34	31.1
Chambery (12)(+27)	16	17.04	30.5
Strasbourg 3 (11)(-6)	17	16.77	30.1
Paris 2 (19)(+4)	18	16.34	29.3
Le Mans (11)(-13)	19	16.10	28.9
Lyon 2 (36)(+13)	20	15.61	28.0
Paris 11 (13)(+20)	21	15.06	27.0
La Rochelle (5)(-8)	22	13.98	25.1
Nancy 2-Strasbourg 1 (59)(0)	23	13.68	24.5
Inra Nancy (6)(+26)	24	12.44	22.3
Paris 12 (7)(+37)	25	11.68	20.9
Nantes (17)(-13)	26	11.28	20.2
Inra Ivry (19)(-1)	27	11.17	20.0
Montpellier 3 (6)(-10)	28	10.89	19.5
Evry (16)(-4)	29	10.78	19.3
Orleans (21)(-2)	30	10.39	18.6
Versailles St Quentin (17)(+27)	31	10.39	18.6
Paris 8 (13)(-3)	32	10.19	18.3
Inra Dijon (8)(-12)	33	10.12	18.2
Mulhouse (6)(-14)	34	9.42	16.9
Perpignan (8)(-23)	35	9.35	16.8
Montpellier 1-Inra (26)(-20)	36	8.95	16.0
Rennes 2 (6)(+16)	37	8.94	16.0
Bordeaux 4 (43)(+7)	38	8.54	15.3
Paris 9 (70)(+13)	39	8.48	15.2
Tours (6)(-13)	40	8.37	15.0
St Etienne (17)(-4)	41	8.30	14.9
Rouen (12)(+21)	42	8.12	14.6
Reims (19)(+4)	43	7.92	14.2

suite page suivante

suite de la page précédente

université	**rg.**	**tot.**	**nor.**
La Rochelle (5)(-1)	44	69.9	1.4
Inra Nancy (6)(+18)	45	68.4	1.3
Angers (14)(+7)	46	66.1	1.3
Montpellier 3 (6)(-3)	47	59.9	1.2
Limoges (9)(+3)	48	57.5	1.1
La Reunion (15)(-8)	49	56.8	1.1
Rennes 2 (6)(+11)	50	53.6	1.0
Littoral (12)(-13)	51	53.5	1.0
Mulhouse (6)(-4)	52	51.8	1.0
Marne La Vallee (10)(-8)	53	51.5	1.0
Lille 3 (6)(+3)	54	46.2	0.9
Poitiers (17)(-4)	55	46.2	0.9
Tours (6)(-4)	56	46.0	0.9
Antilles Guyane (13)(-11)	57	41.4	0.8
Le Havre (11)(+2)	58	40.2	0.8
Metz (7)(-3)	59	37.6	0.7
Toulon (5)(-5)	60	35.9	0.7
Brest (12)(-14)	61	30.6	0.6
Paris 5 (6)(-8)	62	28.2	0.5
Lyon 1 (5)(+2)	63	24.5	0.5
Artois (7)(-5)	64	17.9	0.4
Amiens (11)(-3)	65	16.8	0.3
Corte (6)(+1)	66	14.0	0.3
Bretagne Sud (7)(-3)	67	10.3	0.2

université	**rg.**	**p.c.**	**nor.**
Lille 3 (6)(-4)	44	7.70	13.8
Dijon (38)(+21)	45	7.27	13.0
Toulon (5)(-15)	46	7.17	12.9
Limoges (9)(-9)	47	6.77	12.1
Lille 1-Polytech Lille (111)(0)	48	6.74	12.1
Grenoble 2-Inra (74)(+7)	49	6.51	11.7
Paris 13 (25)(-1)	50	6.50	11.7
Caen-Rennes 1 (89)(-12)	51	6.03	10.8
Metz (7)(-10)	52	5.79	10.4
Marne La Vallee (10)(-18)	53	5.42	9.7
Nice (38)(0)	54	5.31	9.5
Paris 5 (6)(-23)	55	5.12	9.2
Lyon 1 (5)(+2)	56	4.91	8.8
Angers (14)(+3)	57	4.72	8.5
Littoral (12)(-22)	58	4.45	8.0
La Reunion (15)(-8)	59	3.79	6.8
Le Havre (11)(+4)	60	3.66	6.6
Antilles Guyane (13)(-15)	61	3.31	5.9
Poitiers (17)(-1)	62	2.72	4.9
Brest (12)(-19)	63	2.66	4.8
Artois (7)(-9)	64	2.56	4.6
Corte (6)(+2)	65	2.33	4.2
Amiens (11)(-1)	66	1.53	2.7
Bretagne Sud (7)(-8)	67	1.48	2.7

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

TABLE 1.29 – Universités, Moins de 50 ans, Indice G, T=Dégressif

université	**rg.**	**tot.**	**nor.**
Tse-Toulouse 1 (74)(0)	1	72.2	100.0
Pse-Paris 1 (115)(0)	2	69.4	96.2
Hec (61)(+1)	3	66.6	92.2
Aix Marseille 2-3 (61)(-1)	4	46.4	64.3
Paris 10 (46)(+3)	5	41.6	57.6
Nancy 2-Strasbourg 1 (59)(0)	6	40.3	55.8
Ec. Polytechnique (19)(+4)	7	39.4	54.5
Crest-Ensae (30)(-3)	8	39.1	54.2
Lyon 2 (36)(+7)	9	33.9	47.0
Inra Vers-Grig (8)(+10)	10	32.6	45.2
Cergy Pontoise (33)(-4)	11	31.9	44.2
Paris 2 (19)(+9)	12	31.2	43.2
Le Mans (11)(+2)	13	30.4	42.1
Caen-Rennes 1 (89)(-4)	14	30.0	41.5
Lille 1-Polytech Lille (111)(-5)	15	30.0	41.6
Paris 9 (70)(-3)	16	29.2	40.5
Bordeaux 4 (43)(0)	17	25.1	34.8
Inra Rennes (11)(+5)	18	25.1	34.8
Grenoble 2-Inra (74)(-5)	19	24.6	34.1
Cired (7)(+16)	20	24.2	33.6
Clermont 1 (19)(-2)	21	23.8	33.0
Lille 2 (10)(+15)	22	22.9	31.8
Chambery (12)(+19)	23	22.8	31.6
Besancon (16)(+2)	24	22.5	31.1
Dijon (38)(+9)	25	21.3	29.6
Inra Ivry (19)(-3)	26	20.6	28.6

université	**rg.**	**p.c.**	**nor.**
Cired (7)(+16)	1	6.24	100.0
Inra Vers-Grig (8)(+1)	2	6.08	97.4
Tse-Toulouse 1 (74)(-2)	3	5.30	84.8
Inra Rennes (11)(+6)	4	4.61	73.9
Pse-Paris 1 (115)(-1)	5	4.58	73.3
Crest-Ensae (30)(-4)	6	4.53	72.5
Ec. Polytechnique (19)(-2)	7	4.19	67.1
Chambery (12)(+35)	8	3.83	61.3
Paris 10 (46)(+6)	9	3.51	56.2
Cergy Pontoise (33)(-3)	10	3.49	55.9
Le Mans (11)(-5)	11	3.40	54.5
Hec (61)(-3)	12	3.31	53.0
Montpellier 3 (6)(+5)	13	3.18	51.0
Inra Nancy (6)(+36)	14	3.17	50.7
Clermont 1 (19)(+4)	15	3.11	49.9
Paris 11 (13)(+25)	16	3.00	48.1
Aix Marseille 2-3 (61)(-9)	17	2.97	47.5
Perpignan (8)(-6)	18	2.95	47.3
Strasbourg 3 (11)(-7)	19	2.95	47.3
Pau (9)(+14)	20	2.93	47.0
Montpellier 1-Inra (26)(-5)	21	2.84	45.5
Paris 12 (7)(+40)	22	2.81	45.0
Nancy 2-Strasbourg 1 (59)(0)	23	2.68	43.0
Rennes 2 (6)(+29)	24	2.57	41.1
Evry (16)(0)	25	2.50	40.0
La Rochelle (5)(-12)	26	2.48	39.7

suite page suivante

suite de la page précédente

université	rg.	tot.	nor.
Paris 11 (13)(+12)	27	20.5	28.4
Montpellier 1-Inra (26)(-16)	28	20.2	28.0
Nice (38)(-3)	29	20.2	28.0
Nantes (17)(-12)	30	19.7	27.3
Versailles St Quentin (17)(+9)	31	19.0	26.3
Evry (16)(-4)	32	18.4	25.5
Reims (19)(0)	33	18.3	25.4
Pau (9)(+15)	34	17.7	24.6
Orleans (21)(-11)	35	17.3	24.0
Strasbourg 3 (11)(-9)	36	16.7	23.1
St Etienne (17)(-6)	37	15.6	21.6
Angers (14)(+15)	38	15.3	21.2
Paris 13 (25)(-10)	39	14.5	20.2
Rouen (12)(+18)	40	13.1	18.1
Inra Nancy (6)(+22)	41	12.4	17.2
Paris 8 (13)(-9)	42	12.4	17.2
Inra Dijon (8)(-8)	43	12.3	17.1
Paris 12 (7)(+22)	44	12.1	16.8
Perpignan (8)(-15)	45	12.0	16.6
Limoges (9)(+5)	46	11.8	16.4
Mulhouse (6)(+1)	47	11.5	16.0
Poitiers (17)(+2)	48	11.2	15.5
Tours (6)(+3)	49	10.9	15.0
Montpellier 3 (6)(-6)	50	10.4	14.4
La Reunion (15)(-10)	51	10.2	14.2
Marne La Vallee (10)(-6)	52	10.2	14.1
Le Havre (11)(+7)	53	9.9	13.7
Toulon (5)(+1)	54	9.8	13.5
Metz (7)(+1)	55	9.5	13.1
Lille 3 (6)(+1)	56	9.4	13.1
La Rochelle (5)(-14)	57	9.3	12.9
Rennes 2 (6)(+3)	58	9.2	12.8
Antilles Guyane (13)(-13)	59	8.2	11.4
Brest (12)(-12)	60	8.2	11.4
Littoral (12)(-23)	61	7.6	10.5
Paris 5 (6)(-8)	62	7.5	10.4
Artois (7)(-4)	63	5.6	7.8
Corte (6)(+3)	64	4.9	6.8
Lyon 1 (5)(0)	65	4.6	6.4
Bretagne Sud (7)(-2)	66	3.9	5.5
Amiens (11)(-5)	67	3.1	4.3

université	rg.	p.c.	nor.
Lyon 2 (36)(+7)	27	2.48	39.7
Inra Ivry (19)(-2)	28	2.27	36.4
Paris 2 (19)(-6)	29	2.27	36.4
Rouen (12)(+33)	30	2.26	36.2
Orleans (21)(-3)	31	2.25	36.1
Besancon (16)(-8)	32	2.19	35.2
Mulhouse (6)(-13)	33	2.18	34.9
Paris 13 (25)(+16)	34	2.18	34.9
Paris 8 (13)(-4)	35	2.18	35.0
Nantes (17)(-23)	36	2.13	34.2
Lille 2 (10)(-7)	37	2.09	33.5
Inra Dijon (8)(-17)	38	2.04	32.6
St Etienne (17)(-2)	39	2.01	32.3
Toulon (5)(-9)	40	2.00	32.0
Bordeaux 4 (43)(+4)	41	1.97	31.5
Paris 9 (70)(+10)	42	1.93	31.0
Versailles St Quentin (17)(+14)	43	1.92	30.7
Reims (19)(+3)	44	1.85	29.7
Grenoble 2-Inra (74)(+11)	45	1.83	29.4
Caen-Rennes 1 (89)(-7)	46	1.79	28.7
Tours (6)(-19)	47	1.79	28.7
Lille 3 (6)(-8)	48	1.72	27.5
Nice (38)(+5)	49	1.71	27.4
Lille 1-Polytech Lille (111)(-2)	50	1.68	26.9
Metz (7)(-9)	51	1.65	26.4
Marne La Vallee (10)(-17)	52	1.61	25.7
Le Havre (11)(+11)	53	1.56	25.0
Dijon (38)(+12)	54	1.49	23.9
Littoral (12)(-19)	55	1.44	23.1
La Reunion (15)(-5)	56	1.30	20.9
Limoges (9)(-19)	57	1.28	20.4
Paris 5 (6)(-26)	58	1.26	20.3
Antilles Guyane (13)(-13)	59	1.05	16.9
Corte (6)(+7)	60	1.03	16.4
Brest (12)(-17)	61	0.97	15.5
Artois (7)(-7)	62	0.93	14.9
Lyon 1 (5)(-5)	63	0.92	14.7
Angers (14)(-4)	64	0.91	14.6
Bretagne Sud (7)(-6)	65	0.78	12.4
Poitiers (17)(-4)	66	0.78	12.5
Amiens (11)(-2)	67	0.74	11.9

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.B. Classements ne gardant que les 10 ou 30 chercheurs les plus productifs

1.B.1. Centres

TABLE 1.30 – Centres, 10 ou 30 plus productifs, Citations totales, T=Dégressif

centre (top 10)	rg.	p.c.	nor.	centre (top 30)	rg.	p.c.	nor.
Gremaq (Tse-Toulouse 1)(0)	1	605.2	100.0	**Gremaq** (Tse-Toulouse 1)(0)	1	243.3	100.0
Pjse (Pse-Paris 1)(0)	2	405.4	67.0	**Pjse** (Pse-Paris 1)(0)	2	177.4	72.9
Ces (Pse-Paris 1)(+2)	3	205.6	34.0	**Ces** (Pse-Paris 1)(0)	3	104.9	43.1
Non-Grecsta *(Crest-Ensae)*(+5)	4	158.5	26.2	***Centre*** *(Iep Paris)*(+3)	4	101.2	41.6
Greghec (Hec)(-1)	5	158.2	26.2	**Greghec** (Hec)(+1)	5	85.2	35.0
Greqam (Aix Marseille 2-3)(-3)	6	135.1	22.3	**Lerna** (Tse-Toulouse 1)(-1)	6	70.1	28.8
Grecsta (Crest-Ensae)(+1)	7	134.1	22.2	*Greqam* (Aix Marseille 2-3)(-3)	7	61.7	25.4
Lerna (Tse-Toulouse 1)(-1)	8	122.6	20.3	***Non-Grecsta*** *(Crest-Ensae)*(+6)	8	59.2	24.4
Centre *(Iep Paris)*(+3)	9	101.2	16.7	**Grecsta** (Crest-Ensae)(-1)	9	57.0	23.4
Preg (Ec. Polytechnique)(-4)	10	95.0	15.7	***Eco. Pub.*** *(Inra Vers-Grig)*(+5)	10	43.1	17.7
Economix (Paris 10)(+2)	11	80.7	13.3	**Preg** (Ec. Polytechnique)(-2)	11	40.4	16.6
Drm *(Paris 9)*(+14)	12	78.7	13.0	**Economix** (Paris 10)(-1)	12	36.7	15.1
Beta (Nancy 2-Strasb. 1)(-2)	13	75.9	12.6	*Beta* (Nancy 2-Strasb. 1)(-3)	13	36.3	14.9
Lameta (Montpellier 1-Inra)(+1)	14	69.3	11.5	***Drm*** *(Paris 9)*(+25)	14	34.5	14.2
Thema (Cergy Pontoise)(-5)	15	59.6	9.9	*Centre* (Ens Cachan)(-4)	15	34.3	14.1
Eco. Pub. *(Inra Vers-Grig)*(+3)	16	51.4	8.5	***Centre*** *(Cired)*(+5)	16	33.4	13.7
Gredeg *(Nice)*(+8)	17	48.8	8.1	**Smart** (Inra Rennes)(-1)	17	28.6	11.8
Gate *(Lyon 2)*(+6)	18	46.1	7.6	*Thema* (Cergy Pontoise)(-7)	18	26.2	10.8
Cerdi *(Clermont 1)*(+9)	19	45.2	7.5	***Lameta*** *(Montpellier 1-Inra)*(+11)	19	24.6	10.1
Centre *(Cired)*(+15)	20	44.7	7.4	**Centre** (Strasbourg 3)(0)	20	22.0	9.1
Leg *(Dijon)*(+31)	21	40.5	6.7	***Cerdi*** *(Clermont 1)*(+13)	21	20.6	8.5
Centre (Besancon)(-1)	22	38.5	6.4	***Gredeg*** *(Nice)*(+11)	22	20.3	8.4
Cepn *(Paris 13)*(+13)	23	37.4	6.2	**Gate** (Lyon 2)(+2)	23	19.5	8.0
Autre (Tse-Toulouse 1)(-8)	24	36.7	6.1	***Gael*** *(Grenoble 2-Inra)*(+7)	24	18.9	7.8
Centre (Ens Cachan)(-2)	25	34.3	5.7	***Centre*** *(Lille 2)*(+22)	25	18.0	7.4
Gretha *(Bordeaux 4)*(+4)	26	33.5	5.5	***Iredu-Eco*** *(Dijon)*(+59)	26	17.7	7.3
Smart (Inra Rennes)(+2)	27	33.1	5.5	***Non-Cermes*** *(Paris 11)*(+31)	27	17.6	7.2
Crem (Caen-Rennes 1)(-11)	28	32.1	5.3	**Ermes** (Paris 2)(0)	28	17.5	7.2
Centre *(Vers. St Quentin)*(+23)	29	30.9	5.1	*Crem* (Caen-Rennes 1)(-11)	29	17.0	7.0
Non-Gate *(Lyon 2)*(+21)	30	29.9	4.9	**Centre** (Besancon)(-1)	30	16.8	6.9
Clerse-Eco *(Lille 1-Poly. Lille)*(+13)	31	29.7	4.9	*Autre* (Tse-Toulouse 1)(-7)	31	16.4	6.7
Non-Drm (Paris 9)(-13)	32	29.7	4.9	***Leg*** *(Dijon)*(+40)	32	15.8	6.5
Ermes (Paris 2)(-1)	33	28.3	4.7	*Non-Drm* (Paris 9)(-6)	33	14.8	6.1
Centre (Strasbourg 3)(-2)	34	28.0	4.6	***Cepn*** *(Paris 13)*(+12)	34	14.6	6.0
Non-Cermes *(Paris 11)*(+28)	35	27.1	4.5	**Gretha** (Bordeaux 4)(+1)	35	14.4	5.9
Centre (Nantes)(-16)	36	26.8	4.4	*Centre* (La Rochelle)(-17)	36	14.0	5.8
Gael (Grenoble 2-Inra)(+1)	37	26.6	4.4	***Centre*** *(Chambery)*(+26)	37	13.8	5.7
Non-Gretha *(Bordeaux 4)*(+3)	38	26.0	4.3	***Centre*** *(Vers. St Quentin)*(+32)	38	13.3	5.5
Cerag *(Grenoble 2-Inra)*(+21)	39	24.9	4.1	***Lef*** *(Inra Nancy)*(+17)	39	12.9	5.3
Lirhe-Eco *(Tse-Toulouse 1)*(+47)	40	23.6	3.9	*Centre* (Nantes)(-17)	40	12.8	5.3
Centre *(Lille 2)*(+16)	41	23.3	3.9	*Lem* (Lille 1-Poly. Lille)(-3)	41	12.8	5.3
Lem (Lille 1-Poly. Lille)(-20)	42	23.2	3.8	***Lirhe-Eco*** *(Tse-Toulouse 1)*(+50)	42	12.6	5.2
Cermes *(Paris 11)*(+56)	43	21.5	3.6	***Cermes*** *(Paris 11)*(+58)	43	12.4	5.1
Iredu-Eco *(Dijon)*(+50)	44	21.0	3.5	***Centre*** *(Pau)*(+13)	44	12.3	5.0
Leo (Orleans)(-10)	45	21.0	3.5	***Clerse-Eco*** *(Lille 1-Poly. Lille)*(+7)	45	11.7	4.8
Centre *(Chambery)*(+22)	46	20.2	3.3	***Centre*** *(Paris 7)*(+41)	46	11.5	4.7
Centre (Le Mans)(-32)	47	20.2	3.3	***Cerag*** *(Grenoble 2-Inra)*(+34)	47	11.5	4.7
Centre *(Pau)*(+10)	48	20.2	3.3	*Cesaer* (Inra Dijon)(-27)	48	11.4	4.7
Lepii *(Grenoble 2-Inra)*(+31)	49	20.1	3.3	***Non-Gate*** *(Lyon 2)*(+24)	49	11.4	4.7
Autre *(Pse-Paris 1)*(+21)	50	18.3	3.0	*Centre* (Le Mans)(-33)	50	11.2	4.6

suite page suivante

suite de la page précédente

centre (top 10)	**rg.**	**p.c.**	**nor.**	**centre (top 30)**	**rg.**	**p.c.**	**nor.**
Centre (Reims)(-6)	51	17.8	2.9	***Lest-Eco*** *(Aix Marseille 2-3)*(+36)	51	9.8	4.0
Aliss (Inra Ivry)(-25)	52	17.1	2.8	*Aliss* (Inra Ivry)(-26)	52	9.6	3.9
Autre *(Aix Marseille 2-3)*(+6)	53	17.0	2.8	*Centre* (Evry)(-18)	53	9.4	3.9
Centre (Paris 8)(-12)	54	16.7	2.8	***Non-Gretha*** *(Bordeaux 4)*(+5)	54	9.4	3.9
Mona-Tsv *(Inra Ivry)*(+46)	55	16.5	2.7	***Mona-Tsv*** *(Inra Ivry)*(+47)	55	9.2	3.8
Centre (Evry)(-19)	56	16.2	2.7	*Moisa* (Montpellier 1-Inra)(-6)	56	8.6	3.5
Autre (Lille 1-Poly. Lille)(-26)	57	16.1	2.7	***Lepii*** *(Grenoble 2-Inra)*(+35)	57	8.1	3.3
Non-Ermes (Paris 2)(-19)	58	14.4	2.4	*Centre* (Lille 3)(-20)	58	8.0	3.3
Centre (La Rochelle)(-19)	59	14.0	2.3	*Leo* (Orleans)(-11)	59	7.9	3.2
Centre (St Etienne)(-10)	60	13.7	2.3	***Centre*** *(Rennes 2)*(+8)	60	7.7	3.2
Lef *(Inra Nancy)*(+16)	61	12.9	2.1	*Centre* (Montpellier 3)(-19)	61	7.4	3.1
Centre *(Paris 12)*(+12)	62	12.6	2.1	*Centre* (Perpignan)(-29)	62	7.4	3.1
Cesaer (Inra Dijon)(-19)	63	12.6	2.1	*Centre* (Paris 8)(-11)	63	6.9	2.8
Lest-Eco *(Aix Marseille 2-3)*(+29)	64	12.3	2.0	*Non-Ermes* (Paris 2)(-19)	64	6.8	2.8
Centre *(Paris 7)*(+31)	65	11.5	1.9	***Autre*** *(Aix Marseille 2-3)*(+17)	65	6.7	2.8
Non-Crem (Caen-Rennes 1)(-2)	66	10.5	1.7	**Centre** (Reims)(0)	66	6.6	2.7
Autre (Montpellier 1-Inra)(-19)	67	10.4	1.7	*Centre* (St Etienne)(-8)	67	6.6	2.7
Centre *(Rouen)*(+21)	68	10.4	1.7	***Autre*** *(Pse-Paris 1)*(+17)	68	6.4	2.6
Centre (Limoges)(-12)	69	9.6	1.6	***Centre*** *(Valenciennes)*(+10)	69	6.4	2.6
Autre (Grenoble 2-Inra)(-25)	70	9.1	1.5	**Non-Cepn** (Paris 13)(+1)	70	6.3	2.6
Centre (Lille 3)(-17)	71	8.8	1.5	*Centre* (Mulhouse)(-31)	71	6.2	2.5
Centre (La Reunion)(-23)	72	8.6	1.4	*Centre* (Cnam)(-21)	72	6.1	2.5
Moisa (Montpellier 1-Inra)(0)	73	8.6	1.4	*Autre* (Lille 1-Poly. Lille)(-32)	73	5.9	2.4
Centre (Perpignan)(-27)	74	8.5	1.4	*Centre* (Toulon)(-13)	74	5.7	2.4
Centre *(Rennes 2)*(+9)	75	7.7	1.3	***Centre*** *(Toulouse 2)*(+28)	75	5.5	2.3
Non-Beta *(Nancy 2-Strasb. 1)*(+4)	76	7.7	1.3	***Autre*** *(Dijon)*(+19)	76	5.4	2.2
Non-Cepn *(Paris 13)*(+4)	77	7.5	1.3	*Centre* (Limoges)(-9)	77	5.4	2.2
Centre (Montpellier 3)(-9)	78	7.4	1.2	*Non-Gredeg* (Nice)(-13)	78	5.3	2.2
Centre *(Angers)*(+3)	79	7.3	1.2	*Autre* (Montpellier 1-Inra)(-24)	79	5.2	2.1
Centre *(Valenciennes)*(+11)	80	6.4	1.1	*Centre* (Marne La Vallee)(-37)	80	5.1	2.1
Centre (Toulon)(-5)	81	6.3	1.0	*Centre* (Lyon 1)(-12)	81	4.9	2.0
Non-Gredeg (Nice)(-7)	82	6.3	1.1	***Centre*** *(Rouen)*(+15)	82	4.9	2.0
Centre (Mulhouse)(-17)	83	6.2	1.0	*Centre* (Tours)(-38)	83	4.9	2.0
Centre (Cnam)(-11)	84	6.1	1.0	***Centre*** *(Paris 12)*(+6)	84	4.7	2.0
Autre *(Dijon)*(+12)	85	5.9	1.0	*Centre* (La Reunion)(-31)	85	4.6	1.9
Centre (Tours)(-25)	86	5.9	1.0	*Centre* (Littoral)(-37)	86	4.2	1.7
Centre (Poitiers)(-32)	87	5.8	1.0	*Centre* (Angers)(-3)	87	4.1	1.7
Centre *(Toulouse 2)*(+15)	88	5.5	0.9	*Autre* (Grenoble 2-Inra)(-24)	88	4.0	1.7
Centre (Littoral)(-29)	89	5.4	0.9	*Non-Crem* (Caen-Rennes 1)(-6)	89	3.6	1.5
Centre (Marne La Vallee)(-20)	90	5.3	0.9	**Non-Beta** (Nancy 2-Strasb. 1)(0)	90	3.3	1.4
Centre (Brest)(-25)	91	5.0	0.8	*Centre* (Artois)(-15)	91	3.2	1.3
Centre (Ant. Guy.)(-27)	92	4.9	0.8	*Centre* (Paris 5)(-30)	92	3.2	1.3
Centre (Lyon 1)(-7)	93	4.9	0.8	*Centre* (Metz)(-14)	93	3.1	1.3
Centre (Le Havre)(-4)	94	4.2	0.7	*Centre* (Le Havre)(-7)	94	2.8	1.2
Centre (Artois)(-12)	95	4.0	0.7	***Centre*** *(Toulouse 3)*(+5)	95	2.8	1.1
Centre (Metz)(-10)	96	3.9	0.6	*Centre* (Brest)(-22)	96	2.5	1.0
Centre (Paris 5)(-20)	97	3.3	0.6	*Centre* (Ant. Guy.)(-21)	97	2.4	1.0
Centre (Amiens)(-10)	98	2.9	0.5	*Centre* (Poitiers)(-23)	98	2.3	0.9
Centre (Toulouse 3)(+1)	99	2.8	0.5	*Centre* (Corte)(-3)	99	1.8	0.7
Non-Economix (Paris 10)(-5)	100	2.6	0.4	*Non-Economix* (Paris 10)(-3)	100	1.8	0.8
Centre (Corte)(-3)	101	1.8	0.3	*Centre* (Bretagne Sud)(-21)	101	1.5	0.6
Centre (Bretagne Sud)(-10)	102	1.5	0.2	*Centre* (Amiens)(-8)	102	1.4	0.6
Centre (Lyon 3)(-2)	103	0.9	0.2	*Centre* (Lyon 3)(-3)	103	0.9	0.4

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

TABLE 1.31 – Centres, 10 ou 30 productifs, Indice G, T=Dégressif

centre (top 10)	rg.	p.c.	nor.	centre (top 30)	rg.	p.c.	nor.
Pjse (Pse-Paris 1)(+1)	1	21.3	100.0	**Pjse** (Pse-Paris 1)(+1)	1	11.9	100.0
Gremaq (Tse-Toulouse 1)(-1)	2	19.3	90.2	**Gremaq** (Tse-Toulouse 1)(-1)	2	11.2	93.9
Ces (Pse-Paris 1)(+2)	3	14.6	68.3	**Ces** (Pse-Paris 1)(0)	3	9.8	82.4
Greqam (Aix Marseille 2-3)(-1)	4	11.0	51.5	***Centre*** *(Iep Paris)*(+3)	4	9.2	77.4
Lerna (Tse-Toulouse 1)(+2)	5	10.8	50.9	**Lerna** (Tse-Toulouse 1)(0)	5	7.1	59.5
Preg (Ec. Polytechnique)(0)	6	9.7	45.3	**Greqam** (Aix Marseille 2-3)(-2)	6	7.0	58.6
Non-Grecsta (Crest-Ensae)(+2)	7	9.6	45.0	**Greghec** (Hec)(-1)	7	6.4	53.3
Centre *(Iep Paris)*(+4)	8	9.2	43.3	***Economix*** *(Paris 10)*(+3)	8	5.6	47.0
Grecsta (Crest-Ensae)(0)	9	9.2	42.9	***Centre*** *(Cired)*(+12)	9	5.5	46.3
Greghec (Hec)(-6)	10	9.1	42.6	***Eco. Pub.*** *(Inra Vers-Grig)*(+5)	10	5.3	44.8
Economix (Paris 10)(+2)	11	9.0	42.4	**Preg** (Ec. Polytechnique)(-1)	11	5.3	44.5
Beta (Nancy 2-Strasb. 1)(-1)	12	8.3	38.8	**Beta** (Nancy 2-Strasb. 1)(-2)	12	5.2	43.9
Thema (Cergy Pontoise)(-3)	13	7.8	36.6	*Grecsta* (Crest-Ensae)(-5)	13	5.0	41.5
Gredeg *(Nice)*(+11)	14	7.3	34.2	**Smart** (Inra Rennes)(+2)	14	4.9	41.5
Centre *(Cired)*(+20)	15	7.1	33.4	**Non-Grecsta** (Crest-Ensae)(-1)	15	4.7	39.1
Crem (Caen-Rennes 1)(+1)	16	6.9	32.3	**Crem** (Caen-Rennes 1)(+2)	16	4.5	37.6
Cepn *(Paris 13)*(+19)	17	6.7	31.3	*Centre* (Ens Cachan)(-6)	17	4.3	36.5
Drm *(Paris 9)*(+9)	18	6.7	31.2	***Drm*** *(Paris 9)*(+21)	18	4.1	34.5
Gate *(Lyon 2)*(+5)	19	6.6	31.1	*Thema* (Cergy Pontoise)(-7)	19	4.1	34.3
Lameta (Montpellier 1-Inra)(-5)	20	6.4	29.9	***Iredu-Eco*** *(Dijon)*(+65)	20	4.0	33.7
Eco. Pub. (Inra Vers-Grig)(-2)	21	6.3	29.3	***Gredeg*** *(Nice)*(+12)	21	3.8	32.2
Cerdi *(Clermont 1)*(+6)	22	6.1	28.6	***Cerdi*** *(Clermont 1)*(+12)	22	3.6	30.5
Gretha *(Bordeaux 4)*(+7)	23	5.9	27.7	**Gate** (Lyon 2)(+2)	23	3.5	29.0
Smart *(Inra Rennes)*(+5)	24	5.7	26.5	***Lef*** *(Inra Nancy)*(+33)	24	3.5	29.0
Non-Drm (Paris 9)(-7)	25	5.3	25.0	***Cepn*** *(Paris 13)*(+21)	25	3.3	27.3
Non-Gate *(Lyon 2)*(+25)	26	5.2	24.5	***Centre*** *(Chambery)*(+37)	26	3.2	27.0
Clerse-Eco *(Lille 1-Poly. Lille)*(+17)	27	5.1	24.0	*Centre* (Strasbourg 3)(-6)	27	3.2	27.2
Centre *(Vers. St Quentin)*(+24)	28	5.0	23.3	**Non-Drm** (Paris 9)(+1)	28	3.2	26.7
Centre (Besancon)(-8)	29	4.8	22.6	***Gretha*** *(Bordeaux 4)*(+7)	29	3.0	25.4
Leg *(Dijon)*(+23)	30	4.8	22.7	***Lem*** *(Lille 1-Poly. Lille)*(+8)	30	3.0	25.5
Iredu-Eco *(Dijon)*(+63)	31	4.7	22.1	***Cermes*** *(Paris 11)*(+70)	31	2.9	24.3
Lepii *(Grenoble 2-Inra)*(+49)	32	4.7	22.0	**Gael** (Grenoble 2-Inra)(-1)	32	2.8	23.9
Autre (Tse-Toulouse 1)(-17)	33	4.6	21.4	**Lameta** (Montpellier 1-Inra)(-2)	33	2.8	23.6
Centre *(Chambery)*(+35)	34	4.6	21.5	***Lest-Eco*** *(Aix Marseille 2-3)*(+55)	34	2.8	23.1
Cerag *(Grenoble 2-Inra)*(+27)	35	4.6	21.6	***Cerag*** *(Grenoble 2-Inra)*(+45)	35	2.7	22.4
Centre (Nantes)(-16)	36	4.5	21.1	***Moisa*** *(Montpellier 1-Inra)*(+15)	36	2.7	22.4
Lem (Lille 1-Poly. Lille)(-14)	37	4.5	21.3	***Leg*** *(Dijon)*(+35)	37	2.6	21.6
Cermes *(Paris 11)*(+61)	38	4.4	20.7	*Centre* (La Rochelle)(-19)	38	2.5	20.8
Leo (Orleans)(-4)	39	4.4	20.6	***Lirhe-Eco*** *(Tse-Toulouse 1)*(+54)	39	2.5	21.1
Autre *(Pse-Paris 1)*(+31)	40	4.3	20.1	*Centre* (Besancon)(-11)	40	2.4	20.0
Centre (Ens Cachan)(-17)	41	4.3	20.4	*Centre* (Perpignan)(-8)	41	2.4	20.4
Non-Gretha (Bordeaux 4)(+1)	42	4.3	20.2	***Centre*** *(Vers. St Quentin)*(+30)	42	2.4	19.8
Centre (Le Mans)(-29)	43	4.2	19.5	***Clerse-Eco*** *(Lille 1-Poly. Lille)*(+12)	43	2.4	20.0
Autre *(Aix Marseille 2-3)*(+15)	44	4.1	19.3	***Non-Gate*** *(Lyon 2)*(+32)	44	2.4	19.9
Centre (Reims)(+1)	45	4.1	19.4	*Centre* (Le Mans)(-28)	45	2.3	19.5
Lirhe-Eco *(Tse-Toulouse 1)*(+43)	46	4.1	19.1	*Centre* (Montpellier 3)(-3)	46	2.3	19.5
Centre (Strasbourg 3)(-15)	47	4.0	18.9	*Centre* (Nantes)(-22)	47	2.3	19.3
Gael (Grenoble 2-Inra)(-10)	48	3.9	18.3	***Centre*** *(Rennes 2)*(+23)	48	2.3	19.0
Centre *(Pau)*(+7)	49	3.6	16.7	*Ermes* (Paris 2)(-17)	49	2.3	19.5
Centre (Evry)(-13)	50	3.5	16.4	***Non-Cermes*** *(Paris 11)*(+13)	50	2.3	19.4
Lef *(Inra Nancy)*(+27)	51	3.5	16.2	*Centre* (Evry)(-16)	51	2.2	18.5
Non-Cermes *(Paris 11)*(+13)	52	3.5	16.6	***Centre*** *(Pau)*(+6)	52	2.2	18.2
Aliss (Inra Ivry)(-26)	53	3.4	15.9	*Cesaer* (Inra Dijon)(-30)	53	2.2	18.4
Autre (Lille 1-Poly. Lille)(-22)	54	3.4	16.1	***Lepii*** *(Grenoble 2-Inra)*(+41)	54	2.2	18.3
Centre (Paris 8)(-11)	55	3.4	15.9	*Aliss* (Inra Ivry)(-29)	55	2.1	17.5
Ermes (Paris 2)(-21)	56	3.4	16.1	*Autre* (Tse-Toulouse 1)(-31)	56	2.1	17.8
Lest-Eco *(Aix Marseille 2-3)*(+36)	57	3.3	15.6	*Centre* (Lille 3)(-17)	57	2.1	17.8
Centre *(Paris 12)*(+16)	58	3.2	14.8	***Non-Cepn*** *(Paris 13)*(+13)	58	2.0	16.7
Mona-Tsv *(Inra Ivry)*(+43)	59	3.2	15.2	***Autre*** *(Aix Marseille 2-3)*(+23)	59	1.9	15.7
Centre *(Rouen)*(+28)	60	3.0	14.2	*Centre* (Cnam)(-8)	60	1.9	16.1
Centre (St Etienne)(-10)	61	3.0	13.9	***Mona-Tsv*** *(Inra Ivry)*(+43)	61	1.9	16.2

suite page suivante

suite de la page précédente

centre (top 10)	rg.	p.c.	nor.
Autre (Grenoble 2-Inra)(-17)	62	2.9	13.6
Autre (Montpellier 1-Inra)(-15)	63	2.8	12.9
Centre (Perpignan)(-16)	64	2.8	13.1
Non-Crem (Caen-Rennes 1)(+1)	65	2.8	13.2
Moisa *(Montpellier 1-Inra)*(+6)	66	2.7	12.5
Non-Ermes (Paris 2)(-28)	67	2.6	12.4
Centre (La Reunion)(-19)	68	2.5	11.6
Centre (La Rochelle)(-28)	69	2.5	11.6
Cesaer (Inra Dijon)(-27)	70	2.4	11.3
Non-Cepn *(Paris 13)*(+11)	71	2.4	11.2
Centre (Lille 3)(-18)	72	2.3	11.0
Centre (Montpellier 3)(-3)	73	2.3	10.9
Centre *(Rennes 2)*(+12)	74	2.3	10.6
Centre (Lille 2)(-18)	75	2.2	10.2
Centre (Poitiers)(-20)	76	2.2	10.1
Centre (Limoges)(-20)	77	2.1	9.7
Non-Beta (Nancy 2-Strasb. 1)(+2)	78	2.1	10.0
Centre (Cnam)(-6)	79	1.9	9.0
Centre *(Le Havre)*(+11)	80	1.9	8.9
Non-Gredeg (Nice)(-5)	81	1.9	9.0
Centre (Toulon)(-6)	82	1.8	8.3
Centre (Littoral)(-23)	83	1.7	8.1
Centre *(Paris 7)*(+13)	84	1.7	8.1
Centre (Ant. Guy.)(-20)	85	1.6	7.4
Centre (Brest)(-19)	86	1.6	7.4
Centre (Marne La Vallee)(-15)	87	1.6	7.7
Centre (Mulhouse)(-22)	88	1.5	7.0
Centre *(Toulouse 2)*(+15)	89	1.5	7.0
Centre (Tours)(-28)	90	1.5	6.8
Centre (Angers)(-9)	91	1.4	6.6
Centre (Amiens)(-4)	92	1.3	6.2
Autre *(Dijon)*(+4)	93	1.1	5.2
Centre (Metz)(-7)	94	1.1	5.1
Centre (Paris 5)(-18)	95	1.0	4.6
Centre *(Toulouse 3)*(+5)	96	1.0	4.7
Centre (Valenciennes)(-4)	97	1.0	4.8
Centre (Artois)(-15)	98	0.9	4.4
Centre (Lyon 1)(-13)	99	0.9	4.3
Centre (Bretagne Sud)(-8)	100	0.8	3.6
Centre (Corte)(-2)	101	0.8	3.9
Non-Economix (Paris 10)(-5)	102	0.8	3.5
Centre (Lyon 3)(-2)	103	0.4	2.1

centre (top 30)	rg.	p.c.	nor.
Centre *(Reims)*(+4)	62	1.8	14.8
Leo (Orleans)(-14)	63	1.8	15.5
Non-Gretha (Bordeaux 4)(-4)	64	1.8	14.8
Centre (Lille 2)(-18)	65	1.7	14.0
Centre *(Paris 7)*(+22)	66	1.7	14.6
Centre (St Etienne)(-7)	67	1.7	14.5
Autre *(Pse-Paris 1)*(+17)	68	1.6	13.4
Centre (Marne La Vallee)(-25)	69	1.6	13.1
Centre (Paris 8)(-16)	70	1.6	13.6
Centre (Toulon)(-7)	71	1.6	13.6
Non-Gredeg (Nice)(-3)	72	1.6	13.4
Autre (Grenoble 2-Inra)(-9)	73	1.5	12.6
Centre (Mulhouse)(-33)	74	1.5	12.5
Centre *(Toulouse 2)*(+30)	75	1.5	12.5
Autre (Lille 1-Poly. Lille)(-35)	76	1.4	11.7
Autre (Montpellier 1-Inra)(-21)	77	1.4	11.5
Centre (La Reunion)(-22)	78	1.4	11.4
Centre *(Rouen)*(+20)	79	1.4	11.9
Non-Ermes (Paris 2)(-31)	80	1.4	11.6
Centre *(Le Havre)*(+6)	81	1.3	10.7
Centre (Littoral)(-32)	82	1.3	11.2
Centre *(Paris 12)*(+9)	83	1.3	10.8
Centre (Limoges)(-17)	84	1.2	9.9
Centre (Tours)(-41)	85	1.2	10.1
Autre *(Dijon)*(+9)	86	1.0	8.5
Centre *(Toulouse 3)*(+13)	87	1.0	8.4
Centre (Valenciennes)(-8)	88	1.0	8.5
Non-Crem (Caen-Rennes 1)(-3)	89	1.0	8.8
Centre (Lyon 1)(-21)	90	0.9	7.7
Centre (Metz)(-11)	91	0.9	7.4
Centre (Paris 5)(-29)	92	0.9	7.8
Centre (Poitiers)(-15)	93	0.9	7.4
Non-Beta (Nancy 2-Strasb. 1)(0)	94	0.9	7.7
Centre (Angers)(-11)	95	0.8	6.6
Centre (Ant. Guy.)(-19)	96	0.8	6.8
Centre (Brest)(-21)	97	0.8	7.0
Centre (Bretagne Sud)(-15)	98	0.8	6.4
Centre (Corte)(+1)	99	0.8	7.0
Centre (Artois)(-24)	100	0.7	6.3
Centre (Amiens)(-7)	101	0.6	5.3
Non-Economix (Paris 10)(-6)	102	0.5	4.3
Centre (Lyon 3)(-3)	103	0.4	3.8

La colonne "rg." donne le rang, la colonne "tot." donne le score total, "nor." le score normalisé par rapport à celui du premier classé, "p.c." donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

1.B.2. Universités

TABLE 1.32 – Universités, 10 ou 30 aux plus productifs, Citations totales, T=Dégressif

université (top 10)	rg.	p.c.	nor.	université (top 30)	rg.	p.c.	nor.
Tse-Toulouse 1(0)	1	675.6	100.0	**Tse-Toulouse 1**(0)	1	279.8	100.0
Pse-Paris 1(0)	2	468.2	69.3	**Pse-Paris 1**(0)	2	235.4	84.1
Crest-Ensae(+2)	3	230.4	34.1	**Crest-Ensae**(+1)	3	107.7	38.5

suite page suivante

suite de la page précédente

université (top 10)	**rg.**	**p.c.**	**nor.**
Hec(0)	4	158.2	23.4
Aix Marseille 2-3(-2)	5	135.1	20.0
Iep Paris(+4)	6	101.2	15.0
Ec. Polytechnique(-1)	7	95.0	14.1
Paris 9(+1)	8	85.4	12.6
Paris 10(+2)	9	80.7	12.0
Nancy 2-Strasbourg 1(-2)	10	76.1	11.3
Montpellier 1-Inra(+1)	11	73.9	10.9
Cergy Pontoise(-5)	12	59.6	8.8
Lyon 2(+6)	13	59.3	8.8
Dijon(+24)	14	51.8	7.7
Inra Vers-Grig(+2)	15	51.4	7.6
Nice(+6)	16	51.2	7.6
Bordeaux 4(-1)	17	47.3	7.0
Clermont 1(+8)	18	45.2	6.7
Cired(+12)	19	44.7	6.6
Grenoble 2-Inra(+1)	20	43.3	6.4
Lille 1-Polytech Lille(-8)	21	40.8	6.0
Paris 11(+22)	22	40.5	6.0
Paris 13(+6)	23	40.1	5.9
Besancon(-1)	24	38.5	5.7
Paris 2(-5)	25	35.6	5.3
Ens Cachan(-2)	26	34.3	5.1
Caen-Rennes 1(-12)	27	33.4	5.0
Inra Rennes(-1)	28	33.1	4.9
Versailles St Quentin(+12)	29	30.9	4.6
Inra Ivry(-5)	30	29.2	4.3
Strasbourg 3(-1)	31	28.0	4.1
Nantes(-14)	32	26.8	4.0
Lille 2(+13)	33	23.3	3.5
Orleans(-6)	34	21.0	3.1
Chambery(+18)	35	20.2	3.0
Le Mans(-21)	36	20.2	3.0
Pau(+9)	37	20.2	3.0
Reims(-2)	38	17.8	2.6
Paris 8(-5)	39	16.7	2.5
Evry(-8)	40	16.2	2.4
La Rochelle(-8)	41	14.0	2.1
St Etienne(-2)	42	13.7	2.0
Inra Nancy(+16)	43	12.9	1.9
Inra Dijon(-9)	44	12.6	1.9
Paris 12(+13)	45	12.6	1.9
Paris 7(+25)	46	11.5	1.7
Rouen(+19)	47	10.4	1.5
Limoges(-2)	48	9.6	1.4
Lille 3(-7)	49	8.8	1.3
La Reunion(-11)	50	8.6	1.3
Perpignan(-14)	51	8.5	1.3
Rennes 2(+11)	52	7.7	1.1
Montpellier 3(+1)	53	7.4	1.1
Angers(+7)	54	7.3	1.1
Valenciennes(+14)	55	6.4	1.0
Toulon(+2)	56	6.3	0.9
Mulhouse(-6)	57	6.2	0.9
Cnam(-2)	58	6.1	0.9
Tours(-11)	59	5.9	0.9
Poitiers(-17)	60	5.8	0.9
Toulouse 2(+14)	61	5.5	0.8
Littoral(-14)	62	5.4	0.8
Marne La Vallee(-8)	63	5.3	0.8
Brest(-13)	64	5.0	0.7
Antilles Guyane(-15)	65	4.9	0.7
Lyon 1(-1)	66	4.9	0.7

université (top 30)	**rg.**	**p.c.**	**nor.**
Iep Paris(+2)	4	101.2	36.2
Hec(0)	5	85.2	30.5
Aix Marseille 2-3(-3)	6	64.8	23.2
Inra Vers-Grig(+5)	7	43.1	15.4
Paris 9(+6)	8	41.7	14.9
Ec. Polytechnique(-2)	9	40.4	14.5
Nancy 2-Strasbourg 1(-2)	10	37.5	13.4
Paris 10(-2)	11	37.1	13.3
Ens Cachan(-1)	12	34.3	12.3
Cired(+11)	13	33.4	11.9
Montpellier 1-Inra(+4)	14	29.6	10.6
Inra Rennes(0)	15	28.6	10.2
Lyon 2(+4)	16	27.9	10.0
Cergy Pontoise(-7)	17	26.3	9.4
Grenoble 2-Inra(+5)	18	24.8	8.9
Dijon(+27)	19	23.8	8.5
Lille 1-Polytech Lille(-7)	20	23.5	8.4
Bordeaux 4(+1)	21	22.2	7.9
Strasbourg 3(-2)	22	22.0	7.9
Nice(+4)	23	21.9	7.8
Caen-Rennes 1(-8)	24	19.1	6.8
Lille 2(+16)	25	18.0	6.4
Clermont 1(+6)	26	17.9	6.4
Paris 11(+38)	27	17.5	6.3
Besancon(+1)	28	16.8	6.0
Paris 13(+5)	29	16.7	6.0
Paris 2(-2)	30	15.2	5.4
La Rochelle(-12)	31	14.0	5.0
Chambery(+20)	32	13.8	4.9
Versailles St Quentin(+24)	33	13.3	4.8
Inra Nancy(+13)	34	12.9	4.6
Nantes(-9)	35	12.8	4.6
Pau(+12)	36	12.3	4.4
Inra Ivry(-1)	37	11.6	4.1
Paris 7(+29)	38	11.5	4.1
Inra Dijon(-15)	39	11.4	4.1
Le Mans(-23)	40	11.2	4.0
Evry(-10)	41	9.4	3.4
Lille 3(-9)	42	8.0	2.9
Orleans(-3)	43	7.8	2.8
Rennes 2(+11)	44	7.7	2.8
Montpellier 3(-8)	45	7.4	2.7
Perpignan(-15)	46	7.4	2.7
Paris 8(-3)	47	6.9	2.5
Reims(+5)	48	6.6	2.4
St Etienne(+1)	49	6.6	2.4
Valenciennes(+12)	50	6.4	2.3
Mulhouse(-16)	51	6.2	2.2
Cnam(-9)	52	6.1	2.2
Toulon(-3)	53	5.7	2.1
Toulouse 2(+21)	54	5.5	2.0
Limoges(-1)	55	5.4	1.9
Marne La Vallee(-18)	56	5.1	1.8
Lyon 1(-1)	57	4.9	1.8
Rouen(+14)	58	4.9	1.7
Tours(-19)	59	4.9	1.8
Paris 12(+9)	60	4.7	1.7
La Reunion(-16)	61	4.6	1.6
Littoral(-20)	62	4.2	1.5
Angers(+3)	63	4.1	1.5
Artois(-4)	64	3.2	1.1
Paris 5(-14)	65	3.2	1.1
Metz(-3)	66	3.1	1.1

suite page suivante

suite de la page précédente

université (top 10)	**rg.**	**p.c.**	**nor.**	**université (top 30)**	**rg.**	**p.c.**	**nor.**
Le Havre(+1)	67	4.2	0.6	**Le Havre**(0)	67	2.8	1.0
Artois(-6)	68	4.0	0.6	***Toulouse 3***(+6)	68	2.8	1.0
Metz(-4)	69	3.9	0.6	*Brest*(-11)	69	2.5	0.9
Paris 5(-11)	70	3.3	0.5	*Antilles Guyane*(-10)	70	2.4	0.8
Amiens(-5)	71	2.9	0.4	*Poitiers*(-12)	71	2.3	0.8
Toulouse 3(+1)	72	2.8	0.4	**Corte**(-1)	72	1.8	0.6
Corte(-1)	73	1.8	0.3	*Bretagne Sud*(-9)	73	1.5	0.5
Bretagne Sud(-4)	74	1.5	0.2	*Amiens*(-4)	74	1.4	0.5
Lyon 3(-1)	75	0.9	0.1	**Lyon 3**(-1)	75	0.9	0.3

La colonne "rg." donne le rang, la colonne "tot." donne le score total, "nor." le score normalisé par rapport à celui du premier classé, "p.c." donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

TABLE 1.33 – Universités, 10 ou 30 plus productifs, Indice G, T=Dégressif

université (top 10)	**rg.**	**p.c.**	**nor.**	**université (top 30)**	**rg.**	**p.c.**	**nor.**
Pse-Paris 1(+1)	1	23.6	100.0	**Pse-Paris 1**(+1)	1	15.3	100.0
Tse-Toulouse 1(-1)	2	21.5	91.1	**Tse-Toulouse 1**(-1)	2	13.5	88.3
Crest-Ensae(+2)	3	12.3	52.2	***Iep Paris***(+3)	3	9.2	60.3
Aix Marseille 2-3(-1)	4	11.1	47.0	**Crest-Ensae**(0)	4	7.7	50.3
Ec. Polytechnique(+1)	5	9.7	40.9	**Aix Marseille 2-3**(-2)	5	7.6	49.4
Iep Paris(+4)	6	9.2	39.0	**Hec**(-1)	6	6.4	41.5
Hec(-3)	7	9.1	38.4	**Paris 10**(+2)	7	5.7	37.3
Paris 10(+4)	8	9.1	38.4	***Cired***(+16)	8	5.5	36.1
Nancy 2-Strasbourg 1(-1)	9	8.7	36.6	**Nancy 2-Strasbourg 1**(-1)	9	5.4	35.6
Lyon 2(+9)	10	8.1	34.1	*Ec. Polytechnique*(-3)	10	5.3	34.7
Cergy Pontoise(-4)	11	7.8	33.1	**Inra Vers-Grig**(+2)	11	5.3	34.9
Nice(+10)	12	7.7	32.4	**Paris 9**(+2)	12	5.2	34.2
Paris 9(-4)	13	7.4	31.2	**Inra Rennes**(+2)	13	4.9	32.3
Bordeaux 4(+2)	14	7.3	31.0	**Caen-Rennes 1**(+2)	14	4.8	31.6
Cired(+16)	15	7.1	30.2	***Grenoble 2-Inra***(+8)	15	4.7	30.5
Caen-Rennes 1(-1)	16	7.0	29.8	***Lyon 2***(+5)	16	4.7	30.8
Grenoble 2-Inra(+5)	17	7.0	29.5	*Lille 1-Polytech Lille*(-4)	17	4.5	29.6
Paris 13(+13)	18	7.0	29.4	*Ens Cachan*(-7)	18	4.3	28.4
Dijon(+19)	19	6.9	29.1	***Nice***(+8)	19	4.2	27.4
Montpellier 1-Inra(-7)	20	6.9	29.1	**Bordeaux 4**(+2)	20	4.1	26.9
Inra Vers-Grig(-4)	21	6.3	26.4	*Cergy Pontoise*(-10)	21	4.1	26.8
Lille 1-Polytech Lille(-8)	22	6.3	26.6	***Dijon***(+24)	22	4.0	26.4
Clermont 1(+3)	23	6.1	25.8	*Montpellier 1-Inra*(-4)	23	4.0	26.4
Paris 11(+20)	24	5.8	24.4	***Paris 13***(+10)	24	3.8	25.1
Inra Rennes(+2)	25	5.7	23.9	***Inra Nancy***(+22)	25	3.5	22.6
Inra Ivry(-1)	26	5.0	21.2	***Chambery***(+26)	26	3.2	21.0
Versailles St Quentin(+15)	27	5.0	21.0	***Clermont 1***(+6)	27	3.2	20.6
Besancon(-5)	28	4.8	20.4	***Paris 11***(+39)	28	3.2	20.6
Chambery(+24)	29	4.6	19.4	*Strasbourg 3*(-6)	29	3.2	21.2
Nantes(-12)	30	4.5	19.1	***Inra Ivry***(+6)	30	2.5	16.1
Orleans(-3)	31	4.4	18.7	*La Rochelle*(-11)	31	2.5	16.2
Ens Cachan(-8)	32	4.3	18.4	*Besancon*(-3)	32	2.4	15.6
Le Mans(-19)	33	4.2	17.6	*Paris 2*(-4)	33	2.4	15.5
Reims(+2)	34	4.1	17.5	**Perpignan**(-2)	34	2.4	15.9
Paris 2(-15)	35	4.0	16.7	***Versailles St Quentin***(+25)	35	2.4	15.4
Strasbourg 3(-5)	36	4.0	17.1	*Le Mans*(-19)	36	2.3	15.2
Pau(+7)	37	3.6	15.1	**Montpellier 3**(+1)	37	2.3	15.2
Evry(-6)	38	3.5	14.8	*Nantes*(-10)	38	2.3	15.0
Inra Nancy(+21)	39	3.5	14.7	***Rennes 2***(+19)	39	2.3	14.8

suite page suivante

suite de la page précédente

université (top 10)	**rg.**	**p.c.**	**nor.**	**université (top 30)**	**rg.**	**p.c.**	**nor.**
Paris 8(-6)	40	3.4	14.4	*Evry*(-9)	40	2.2	14.4
Paris 12(+16)	41	3.2	13.4	*Inra Dijon*(-16)	41	2.2	14.4
Rouen(+24)	42	3.0	12.8	***Pau***(+8)	42	2.2	14.2
St Etienne(-2)	43	3.0	12.5	*Lille 3*(-10)	43	2.1	13.9
Perpignan(-7)	44	2.8	11.8	**Cnam**(-1)	44	1.9	12.6
La Reunion(-6)	45	2.5	10.5	*Orleans*(-4)	45	1.9	12.1
La Rochelle(-12)	46	2.5	10.5	***Reims***(+7)	46	1.8	11.5
Inra Dijon(-12)	47	2.4	10.2	*Lille 2*(-6)	47	1.7	10.9
Lille 3(-6)	48	2.3	9.9	***Paris 7***(+20)	48	1.7	11.3
Montpellier 3(+6)	49	2.3	9.8	**St Etienne**(+2)	49	1.7	11.3
Rennes 2(+15)	50	2.3	9.6	*Marne La Vallee*(-12)	50	1.6	10.2
Lille 2(-5)	51	2.2	9.2	*Paris 8*(-6)	51	1.6	10.6
Poitiers(-8)	52	2.2	9.1	**Toulon**(0)	52	1.6	10.6
Limoges(-7)	53	2.1	8.8	*Mulhouse*(-18)	53	1.5	9.7
Cnam(+2)	54	1.9	8.1	***Toulouse 2***(+22)	54	1.5	9.8
Le Havre(+14)	55	1.9	8.0	*La Reunion*(-10)	55	1.4	8.9
Toulon(+2)	56	1.8	7.5	***Rouen***(+16)	56	1.4	9.3
Littoral(-9)	57	1.7	7.3	***Le Havre***(+10)	57	1.3	8.4
Paris 7(+14)	58	1.7	7.3	*Littoral*(-15)	58	1.3	8.7
Antilles Guyane(-9)	59	1.6	6.7	***Paris 12***(+12)	59	1.3	8.4
Brest(-8)	60	1.6	6.7	*Limoges*(-6)	60	1.2	7.7
Marne La Vallee(-4)	61	1.6	6.9	*Tours*(-22)	61	1.2	7.9
Mulhouse(-11)	62	1.5	6.3	***Toulouse 3***(+11)	62	1.0	6.6
Toulouse 2(+13)	63	1.5	6.3	**Valenciennes**(0)	63	1.0	6.6
Tours(-14)	64	1.5	6.1	*Lyon 1*(-8)	64	0.9	6.0
Angers(-4)	65	1.4	5.9	**Metz**(-1)	65	0.9	5.7
Amiens(0)	66	1.3	5.6	*Paris 5*(-14)	66	0.9	6.1
Metz(-2)	67	1.1	4.6	*Poitiers*(-5)	67	0.9	5.8
Paris 5(-9)	68	1.0	4.1	**Angers**(-2)	68	0.8	5.1
Toulouse 3(+5)	69	1.0	4.3	*Antilles Guyane*(-8)	69	0.8	5.3
Valenciennes(+1)	70	1.0	4.3	*Brest*(-10)	70	0.8	5.4
Artois(-9)	71	0.9	4.0	*Bretagne Sud*(-4)	71	0.8	5.0
Lyon 1(-7)	72	0.9	3.9	***Corte***(+3)	72	0.8	5.4
Bretagne Sud(-3)	73	0.8	3.3	*Artois*(-13)	73	0.7	4.9
Corte(-1)	74	0.8	3.5	*Amiens*(-4)	74	0.6	4.1
Lyon 3(-1)	75	0.4	1.9	**Lyon 3**(-1)	75	0.4	2.9

La colonne “rg.” donne le rang, la colonne “tot.” donne le score total, “nor.” le score normalisé par rapport à celui du premier classé, “p.c.” donne le score par chercheur. Entre les premières parenthèses se trouve le nombre de chercheurs, la variation de classement par rapport à Clm est donnée entre les deuxièmes parenthèses. Si cette variation est supérieure ou égale à 3, le nom est typographié en gras italique, si elle est inférieure ou égale à -3, le nom est typographié en italique, sinon le nom est simplement typographié en gras.

Chapter 2

Are Researchers who Publish More also More Cited? Individual Determinants of Publications and Citations Records [1]

Thanks to an unique individual dataset of French researchers in economics, we explain individual publication and citation records with individual demographic characteristics (gender and age), coauthorship patterns (average number of authors per article and size of coauthors' network) and specialisation choices (percentage of output in each JEL codes). The analysis is performed on both EconLit publication scores (adjusted for journals quality) and Google Scholar citation indexes, which allows us to present a broad picture of

1. This chapter has been co-written with Pierre-Philippe Combes. It has benefited from the excellent research assistance of Philippe Donnay and Charles Laitong to whom we are very grateful.

knowledge diffusion in economics. Citations are largely driven by publication records but also importantly increased by larger research team sizes and coauthors' networks.

2.1. Introduction

The objectives of any academic research activity can be numerous: answering a specific question, understanding some mechanisms, evaluating a public policy, replicating other studies for robustness tests, etc. Also, academic researchers are often interested in the diffusion of their findings through publications. A common tool to measure this knowledge diffusion consists in using citations from other researchers. These citations can be collected in other journal articles, but also in books, working papers, etc. We focus here on publications as journal articles referenced in the EconLit database and citations collected by Google Scholar of an exhaustive census of French academic economists in 2008. The attention is made about the determinants of these records which we consider as measures of research productivity.

We propose to answer three sets of questions. First, are the standard determinants of productivity also determinants of productivity in research? Second, are publications and citations records driven by field choices? Third, to what extent journal quality weights and citations measure the same dimension of publication productivity and, when it is not the case, how is knowledge diffused among researchers? We use an exhaustive dataset of French academic economists in 2008, their quality-adjusted publication records in EconLit and their Google Scholar citation indexes. Next to the researcher's age, age squared and gender, we first introduce a specific variable related to the organisation of labour in the academic research activity, which is the average number of authors per publication. We find that larger research teams have a more robust impact than standard arguments in a Mincerian equation like age or gender in explaining publication scores and citation indexes. We also introduce the quantity of published articles and coauthors' network size (total number of different coauthors) as determinants of the average quality of publications and find increasing returns to scale: researchers who have published more articles and who have had more different coauthors reach higher average quality of publications.

Then, we introduce specialisation patterns measured with the shares of researchers' journal articles in each JEL (Journal of Economic Literature) classification code. It turns out that, even if these specialisation choices do not considerably impact the overall diagnosis described above, we find some evidence of field disparities, both in terms of publications' and citations' patterns. For instance, French researchers in economics specialised in the fields of Microeconomics and Labor and Demographic Economics publish more articles, of a higher average quality, hence reach higher total publication scores and are more cited than the average.

Finally, regressing citation indexes on publication scores and other individual characteristics, we find that the most important part of the variance of citations is explained with publication outputs, which allows us to conclude that journal quality weights and citations globally measure the same dimension of the publication productivity. Nevertheless, we observe some deviation for specific over- or under-cited fields and strong network effects related to the organisation of the research and publication activity. At given publication records, discounted by the number of authors per article, larger team sizes and larger coauthorship networks generate more citations. We interpret this result as a decomposition between a direct and an indirect effect. The former stands because the different coauthors of an article present in different conferences, seminars, informal talks, etc. The latter stands because researchers talk about their new papers also to their former coauthors, which in turn diffuse the knowledge to their colleagues and other coauthors.

In the 1970s, the quantity of scholars' publications has been proved to be an important determinant of academic wages as it is used to evaluate an reward university professors. Katz (1973) evaluates the return to an article between \$18 and \$102 in 1969 in one large US public university. In a multi-equation system modelling job-quality, research productivity and earnings of a 863-economists sample in 1966, Hansen, Weisbrod and Strauss (1978) find that experience, measured by the number of years since Ph.D., has a significant impact on research productivity. They also evaluate the return to an additional unit of research productivity to an almost 8% increase in annual earnings.

From the 1980s, the quality of publications, measured by citations, has been proved to be a much more important determinant of salaries than the quantity of publications. On a sample of 148 full professors of economics at 7 large public universities, Hamermesh, Johnson and Weisbrod (1982) show that citations, which they define as indirect contributions to knowledge, have a superior effect on academic earnings than the number of publications, which they define as direct contributions to knowledge. Diamond (1986) goes further in the analysis and estimates that the marginal value of a citation lies between \$50 and \$1,300 depending on disciplines. With a panel of 140 academic economists, Sauer (1988) confirms this existence of incentives able to promote knowledge growth and estimates that an individual's return from a coauthored paper with n authors is approximately $1/n$ times that of a single-authored paper. Finally, Kenny and Studley (1995) show that economists' salaries are best characterised by implicit long-term contracts (predicted publications and citations) than by expected near-term productivity (presumably because of mobility cost) and find an insignificant effect of field choice on academic wages.

Adopting a macro perspective, Lovell (1973) estimates production functions of publications and citations, that he defines as contributions to economics knowledge. At the aggregate level, considering previously published articles as the stock of capital and the number of granted Ph.D. in the USA as the labour input in a Cobb-Douglas production function, he explains the tendency of scientific literature to grow exponentially between 1895 and 1965 by the exponential growth of the labour input. Here, we take another road and study the micro determinants of publications and citations which have been proved to be important determinants of academic wages and promotions. Hence, this should allow us to shed some light on optimal behaviour of academic researchers concerning their collaborations, their specialisation choices and more generally their research strategy. From a policy point of view, for the efficiency of the academic system, it is important to understand to which incentives researchers best react in this activity. This could also rise another question. What happens in a country in which incentives are less directly related to publication? In our sample of French economists, we do not observe wages to properly answer this question and this is hence left for further research.

Finally, Stigler and Friedland (1975) is the most closely article related to ours. They examine the citations of articles published between 1950 and 1968 in two economic subfields by doctorates in economics from six major US universities to identify intellectual debtors' and creditors' patterns. Regressing the number of citations, considered as a measure of influence on the number of published articles, they find some evidence of weak but significant increasing returns to quantity.

To our knowledge, we are the first to regress citation indexes on publication outputs taking quality into account individual characteristics (including coauthorship patterns) and specialisation choices. Another contribution of ours is to use a new tool to measure the impact of a researcher, its Google Scholar citations. Standard studies on citations use the Thomson Reuters (JCR) database but it refers only 304 journals in economics. Hence, both citations of articles in non-referenced journals and citations from articles in non-referenced journals are excluded. Using Google Scholar citations has the decisive advantage to take into account a much larger support regarding cited and citing articles, journals but also books, working papers, policy reports, etc., since all supports on the internet on academic web sites are concerned.

The rest of the paper is organised as follows. Data and econometric strategy are presented in sections 2.2 and 2.3, respectively. Individual determinants of publication scores and citation indexes are analysed in section 2.4. Section 2.5 tests the robustness of the findings when specialisation choices are taken into account. Comparing citation indexes and publication scores, section 2.6 analyses the patterns of knowledge diffusion while section 2.7 concludes.

2.2. Data

2.2.1. Measure of output

We measure the research output of a researcher i in two ways, her journal quality adjusted number of publications and her number of citations in Google Scholar.

2.2.1.1. Publication records

Publication records are measured as weighted sums of publications. All publications come from the EconLit database listing more than 560,000 papers published in more than 1200 journals between 1969 and 2008. Three dimensions enter the weighted scheme of publications, the relative number of pages, the number of authors and the quality of journals.

We account for the number of pages of articles to capture the idea that longer articles contain more ideas (normal vs short papers like notes in the *American Economic Review*, for instance). Nevertheless, since the layout can be very different from a journal to another and since we do not want to favorise some journals, the weighting is made within each journal in the sense that the ratio of the number of pages of article a over the average number of pages of articles published in the journal is used as weight. Since editor's policy can vary over time, the average number of pages is calculated for each year separately.

We also account for the number of authors of publications. For aggregational purpose, we assume an equal split of the publication between its authors. [2] This allows us to test for the presence of increasing, or decreasing, returns to scale within authors' teams by using the average number of authors as an independent variable.

Finally, we take the quality of publications into account using Combes and Linnemer (2010) journals weighting schemes. For robustness purposes, we compare our results using two of their indexes assuming different degrees of convexity in the distribution of journals'

2. As a consequence, to evaluate the research output of an academic department, an article written by two members of the same department would count as much as the same article written by only one author.

weights. To sum up, the output of researcher i is a weighted sum of her articles a:

$$y_i = \sum_a \frac{W(a)}{n(a)} \frac{p(a)}{\bar{p}} \tag{2.1}$$

where $p(a)$ is the number of pages of the article, $\bar{p}$ the annual average number of pages of articles in the journal, $n(a)$ the number of authors of the article and $W(a)$ the weighting scheme of journals. We consider the medium and high degree of convexity of journals' weights, noted CLm and CLh respectively. Scores using neither journals weights nor the correction for the relative number of pages are noted $E1n_a = 1/n(a)$. CLm goes from a weight equal to 100 for the *Quarterly Journal of Economics* to a weight of 4 for the last journal, passing by 55.1 for the *Journal of Labor Economics* for instance. CLh goes from 100 to the *Quarterly Journal of Economics* to 0.0007 for the last journal, passing by 16.7 for the *Journal of Labor Economics*. We refer to these two schemes as the "Quality" and "Top quality" publication measures respectively. They are illustrated for the top 50 journals in Table 3.16 in Appendix 3.A. $E1n$ is referred to as "Quantity".

2.2.1.2. Citation records

We assess citations through the Google Scholar citations of articles, books and working papers written by the researchers of our database. These citations were extracted on January 2010, around two years after the date at which we want to measure research impacts, which seems quite reasonable given the time needed for works to be cited.

In order to avoid homonyms problems for researchers with names identical to ours in fields other than economics, we restrict fields on Google Scholar to the "subject areas"

"Business, Administration, Finance, and Economics" and "Social Sciences, Arts, and Humanities". To have a period of time comparable to the one we use for EconLit, we keep only works published between 1969 and 2008.

Like for publications, we account for the number of authors for each work. However, it is no more necessary to account for the support's quality or for the relative length of publications since we assume that the number of citations directly reflect works' quality.

We first built an index of total citations, $TCit_i$, which is the total number of citations received by all researcher i's works discounted by the number of authors. Then, to combine this total volume with something closer to the average quality of publications, we use a synthetic index. We choose not to use the famous H-index, noted H_i, proposed by Hirsh (2005) because typically two researchers can have the same H-index when one of them has some highly cited publications and the other not. In other words, the H-index ignores the internal distribution of citations received by the articles used to calculate it. Therefore we prefer to use the G-index, proposed by Egghe (2006) which stands that researcher i has a G_i-index equal to g, which is unique, if her g most cited articles have received, together, g^2 citations, or g citations on average. It can be shown that $G_i \geq H_i$. The difference between the two indexes is relates to the number of citations received on average by the most cited articles.

Both H and G indexes ignore the bottom of the distribution of citations but the G-index has the advantage to better take account to publications. To take the number of authors of articles into account, we follow Schreiber (2008) who proposes to attribute all citations to the article but simply a fraction of the article to each author. Then H and G indexes are not necessary integers anymore but they keep the same signification. For instance, a G-index of 7.5 means that the researcher has published at least 7.5 articles equivalent written alone with 7.5 citations each on average.

2.2.2. Population and descriptive statistics

The EconLit database, which enables us to compute our different scores of publication, is merged with a list of researchers in economics given by the French Ministry of Education and Research. [3] In 2008, year at which the analysis is conducted, there are 2782 researchers. [4]

Table 2.1 presents various descriptive statistics about our main dependent and independent variables. The average researcher is around 47 years old, has published 3.5 articles (equivalent alone), has been cited 114 times and has a G-index of 7.71 (which means that her 7.71 most cited papers have been cited 7.71 times on average). In 2008, 30% of French academic economists are women and 73% have published at least one article in the EconLit database between 1969 and 2008. We define them as "Publishers". Also, 85% have been cited at least once on Google Scholar; we define them as "Cited".

If we restrict our sample to researchers who have published at least one article referenced in the EconLit database (Panel (b)), the average publisher is slightly younger (around 46 years old) and more likely to be a man (27% against 30% of women for all researchers). She has published 4.8 articles (equivalent alone), has been cited 139 times and has a G-index of 9 on average. 95% of publishers have been cited at least once on Google Scholar. The average number of authors per article is around 2 (1.85 in EconLit, 2.05 in Google Scholar)

3. Ministère de l'Enseignement Supérieur et de la Recherche - Direction Générale de la Recherche et de l'Innovation.

4. We had to merge data from the university system, from the CNRS (Centre National de la Recherche Scientifique) and from the INRA (Institut National de la Recherche Agronomique).

Table 2.1: Individual descriptive statistics

	Observations	Minimum	Maximum	Mean	Stand. err.
Panel (a): All academics					
Women	2782	0	1	0.30	0.46
Age	2782	26	85	46.87	10.79
Publisher	2782	0	1	0.73	0.44
Quantity	2782	0	78.5	3.53	6.07
Quality total score	2782	0	5866.6	59.99	209.57
Top quality total score	2782	0	3867.7	16.46	112.90
Cited	2782	0	1	0.85	0.35
Total citations	2616	0	30069.0	113.97	672.00
G-index	2616	0	165.8	7.71	10.46
GS Authors number	2616	1	6	2.02	0.63
Panel (b): Publishers					
Women	2040	0	1	0.27	0.44
Age	2040	27	85	46.09	10.58
Quantity	2040	0.17	78.5	4.81	6.64
Quality total score	2040	0.40	5866.6	81.81	241.08
Top quality total score	2040	0	3867.7	22.45	131.34
Average quality	2040	0.65	114.2	12.03	11.20
Average top quality	2040	0	73.0	1.84	5.69
EL Authors number	2040	1	6	1.85	0.63
Network size	2040	0	53	4.12	5.35
Cited	2040	0	1	0.95	0.23
Total citations	2015	0	30069.0	139.19	762.57
G-index	2015	0	165.8	9.00	11.31
GS Authors number	2015	1	5	2.05	0.57
Panel (c): Cited					
Women	2374	0	1	0.28	0.45
Age	2374	27	85	46.31	10.70
Publisher	2374	0	1	0.81	0.39
Quantity	2374	0	78.5	4.05	6.41
Quality total score	2374	0	5866.6	69.75	225.38
Top quality total score	2374	0	3867.7	19.27	122.00
Average quality	1929	0.65	114.2	12.35	11.41
Average top quality	1929	0	73.0	1.94	5.84
EL Authors number	1929	1	5	1.87	0.61
Network size	2374	0	53	3.49	5.17
Total citations	2374	0.20	30069.0	125.59	704.39
G-index	2374	0.20	165.8	8.50	10.68
GS Authors number	2374	1	5	2.05	0.58

Quantity = $E1n$, number of articles equivalent written alone in EconLit. Quality and top quality total scores = CLm and CLh, publication scores with a low and high degree of convexity in the journals weighting scheme, respectively. Authors number = average number of authors by article. Network size = total number of different coauthors.

while the average network size (total number of different coauthors) is 4.12.

Alternatively, if we restrict our sample to researchers who have been cited at least once on Google Scholar (Panel (c)), 81% have published at least one article in EconLit and 28% are women. The average cited researcher has published 4.05 articles (equivalent alone) in EconLit, has been cited 126 times and has a G-index of 8.5. The average network size of cited researchers is slightly smaller (3.49 vs 4.12) than that of publishers.

Finally, Table 2.2 provides some simple correlations between EconLit publication scores and Google Scholar citation indexes.[5] Notable facts are the following. As expected, researchers who have published more articles are more cited. The correlation between citation indexes and publication scores is higher when the quality of journals is taken into account with a medium degree of convexity in the journals weighting scheme than with a high degree of convexity. But the average quality of publications is more correlated with citation indexes when there is a high degree of convexity in the journals weighting scheme.

Table 2.2: Correlations of EconLit and Google Scholar indexes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Quantity (1)	1	0.89	0.70	0.32	0.43	0.61	0.60
Quality total score (2)		1	0.93	0.72	0.77	0.64	0.64
Top quality total score (3)			1	0.85	0.94	0.59	0.60
Average quality (4)				1	0.93	0.41	0.40
Average top quality (5)					1	0.47	0.48
Total citations (6)						1	0.95
G-index (7)							1

Quantity = $E1n$, number of articles equivalent written alone in EconLit. Quality and top quality total scores = CLm and CLh, publication scores with a low and high degree of convexity in the journals weighting scheme, respectively. Average quality and top quality = $CLm/E1n$ and $CLh/E1n$. These correlations have been calculated on logarithms of the scores to follow the econometric specification presented in section 2.3.

5. In logarithms to follow the econometric analysis presented in section 2.3.

2.3. Econometric specification

For our different measures of publication scores or citation indexes, we estimate the following specification using ordinary least squares:

$$log\ y_i = \beta_0 + \beta_1\ Gender_i + \beta_2\ Age_i + \beta_3\ Age_i^2 + \beta_4\ log\ \overline{nau}_i + \beta_5\ log(1 + Net_i) + \sum_{j=1}^{18} \gamma_j \frac{y_{ij}}{y_i} + \epsilon_i$$

where y_i is the publication score or citation index of researcher i, $Gender_i$ is a dummy variable taking 1 for women, $\overline{nau}_i$ is the average number of authors per publication of researcher i and Net_i is the size of its coauthorship network, *i.e.* its total number of different coauthors.[6] Finally, $\frac{y_{ij}}{y_i}$ is the share of researcher i's output in each JEL codes j at the first letter level (calculated in terms of $E1n$, the number of articles published in EconLit, equivalent written alone).[7] [8]

Since some researchers have never published any article in EconLit or have no citation on Google Scholar, we also take selection into account using a Heckman 2-step procedure. The first step is the selection probit equation and the second step is the main equation augmented by the inverse of Mills' ratio. In that case, this corresponds to a model where researchers who have not published or are not cited are those who do not reach a sufficient quality threshold in their research activity and this can be explained by the same type of variables as those who determine the level of publication or citations. Unfortunately, for

6. The researchers who have never coauthored any article have a network size of size 0, so we add 1 to Net_i to take its log. $1 + Net_i$ can be seen as the total network size, including researcher i who would belong to its own network in this case.

7. We ignore the fields "Miscellaneous Categories" (Y) and "Other Special Topics" (Z). We also slightly modify the codes C and D by merging code C7 (Game Theory and Bargaining Theory) and C9 (Design of Experiments) to Microeconomics (D), which seems to us more consistent.

8. To estimate all γ_j coefficients, we constraint the regressions with $\sum_{j=1}^{18} \gamma_j = 0$.

that very reason, it is pretty difficult to find exclusion restrictions, variables that would explain the probability to publish or be cited, but not the levels. Therefore the selection effect is identified on non-linearities only.

2.4. Individual determinants of publication and citation records

We analyse in this section the individual determinants of publication and citation records regressing total publication scores ($E1n$, number of articles equivalent written alone, CLm and CLh, quality and top quality publication scores using a medium and a high degree of convexity in the journals weighting scheme, respectively) and citation indexes ($TCit$, Google Scholar total citations discounted by the number of authors per paper and G, Google Scholar G-index) on gender, age and its square and the average number of authors per article. This is present in Table 2.3. We also regress $CLm/E1n$ and $CLh/E1n$, the average quality and top quality of publications on the same explicative variables augmented by $E1n$ (quantity) and the network size to identify possible increasing returns to quantity and coauthorship. Since network size is by construction highly correlated with quantity, it is included in the regressions only when quantity is also included to identify network effects on top of its natural quantity effect. Except for age and gender, all variables are in logs.

Women are less productive, whatever the mesure of research output. Older researchers publish more articles, which are on average of lower quality. These two effects cancel out each other when the dependent variable is the total publication score (taking quality of publications into account). Olders are also more cited (with a slightly concave effect), which could be due to the fact that their articles were published less recently and hence, had more time to be cited. Moreover, we find increasing returns to the average number of authors per article for all research output measures, except for the number of articles equivalent written alone. A publisher that has on average two coauthors instead of one only (per article) has 10.1% less publications but their average quality is 8.4% higher and their average top quality is 36.8% higher. [9] Her total quality and top quality publication scores are 11.4% and 96.9% higher, respectively, whereas she is cited 53.4% more and has a G-index 41.8% higher. [10]

We also find increasing returns to individual quantity and size of the coauthorship network on average quality of publications. The more researchers publish, the more different coauthors they have had, the higher the average quality of their publications. Researches who have published five articles equivalent written alone instead of four have an average quality of publications 3% higher and an average top quality of publications 15.5% higher. [11] Having a stock of five different coauthors instead of four, meaning a total network size of 6 instead of 5, increases average publication quality by 4% and average publication top quality by 15.5%. [12]

In Table 2.4, we repeat the same exercice as in Table 2.3 except that we take selection into account using a Heckman 2-step procedure. Hence, in the first column of Table 2.4, we run a probit equation of the probability to have published, which allows us to calculate the inverse of Mills' ratio ("Selection") that we include in the columns 2 to 6. The same exercice is done for Google Scholar citation indexes in columns 7 to 9. Without clear exclusion restrictions, the inverse of Mills' ratio should at least control for the presence of

9. $1.5^{-0.262} - 1$, $1.5^{0.200} - 1$ and $1.5^{0.773} - 1$, respectively.
10. $1.5^{0.266} - 1$, $1.5^{1.671} - 1$, $1.5^{1.056} - 1$ and $1.5^{0.861} - 1$, respectively.
11. $1.25^{0.132} - 1$ and $1.25^{0.644} - 1$, respectively.
12. $1.2^{0.216} - 1$ and $1.2^{0.789} - 1$, respectively.

Table 2.3: Determinants of publication scores and citation indexes: individual characteristics

	E1n	CLm	CLh	CLm/E1n	CLh/E1n	TCit	G
Women	-0.357[a]	-0.556[a]	-1.126[a]	-0.106[a]	-0.372[a]	-0.517[a]	-0.298[a]
	(0.05)	(0.07)	(0.13)	(0.03)	(0.09)	(0.08)	(0.04)
Age	0.059[b]	0.029	-0.037	-0.048[a]	-0.171[a]	0.104[a]	0.055[b]
	(0.03)	(0.04)	(0.07)	(0.01)	(0.04)	(0.04)	(0.02)
Age^2	-0.000[c]	-0.000	0.000	0.000[a]	0.001[a]	-0.001[c]	-0.000[c]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number	-0.262[a]	0.266[a]	1.671[a]	0.200[b]	0.773[a]	1.056[a]	0.861[a]
	(0.07)	(0.10)	(0.18)	(0.09)	(0.28)	(0.12)	(0.08)
Quantity				0.132[a]	0.644[a]		
				(0.03)	(0.08)		
Network size				0.216[a]	0.789[a]		
				(0.04)	(0.14)		
Constant	-0.386	2.501[a]	-0.504	3.134[a]	0.848	-0.535	-0.486
	(0.59)	(0.82)	(1.65)	(0.31)	(0.95)	(0.87)	(0.48)
R^2	0.059	0.034	0.066	0.251	0.355	0.082	0.101
Observations	2040	2040	2040	2040	2040	2374	2374

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. $E1n$ = number of articles equivalent written alone in EconLit. CLm and CLh are publication scores with a low and high degree of convexity in the journals weighting scheme, respectively. $CLm/E1n$ and $CLh/E1n$ = average quality and top quality. $TCit$ = Google Scholar total citations ; G = Google Scholar G-index.

some non-linearities in the model.

Comparing results from Table 2.3 and Table 2.4, we find that if women have published less articles in EconLit on average, it is because of the women who have not published at all. Once controlling for this selection process, publishing women have published as many papers as men and cited women are as much cited than cited men.

Taking selection into account, older researchers do not produce lower quality articles but they are still more cited. Results on the increasing returns to the average number of authors per articles and to the quantity of publications and the size of the coauthorship network for the average quality of publications are not impacted by the Heckman procedure.

Table 2.4: Determinants of publication scores and citation indexes, with selection

	Prob.Pub	E1n	CLm	CLh	CLm/E1n	CLh/E1n	Prob.Cit	TCit	G
Women	-0.405^a	-0.781	-1.385	-3.580^b	-0.477^c	-2.215^a	-0.325^a	-0.408	-0.083
	(0.06)	(0.65)	(0.87)	(1.63)	(0.27)	(0.79)	(0.07)	(0.71)	(0.42)
Age	0.028	0.096^c	0.103	0.181	-0.015	-0.008	0.006	0.100^b	0.048^b
	(0.03)	(0.05)	(0.07)	(0.14)	(0.03)	(0.08)	(0.03)	(0.04)	(0.02)
Age^2	-0.000^c	-0.001	-0.001	-0.003	-0.000	-0.001	-0.000	-0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number		-0.262^a	0.267^a	1.673^a	0.195^b	0.747^a		1.056^a	0.861^a
		(0.07)	(0.10)	(0.18)	(0.09)	(0.28)		(0.12)	(0.08)
Quantity					0.130^a	0.634^a			
					(0.03)	(0.08)			
Network size					0.219^a	0.805^a			
					(0.04)	(0.14)			
Selection		2.204	4.300	12.736	1.925	9.565^b		-1.251	-2.465
		(3.38)	(4.51)	(8.42)	(1.39)	(3.95)		(8.31)	(4.93)
Constant	0.571	-1.625	0.084	-7.662^c	2.054^b	-4.517^c	1.628^b	-0.367	-0.155
	(0.59)	(1.76)	(2.36)	(4.49)	(0.85)	(2.44)	(0.77)	(1.27)	(0.75)
R^2		0.060	0.035	0.068	0.251	0.356		0.082	0.101
Observations	2782	2040	2040	2040	2040	2040	2782	2374	2374

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. $E1n$ = number of articles equivalent written alone in EconLit. CLm and CLh are publication scores with a low and high degree of convexity in the journals weighting scheme, respectively. $CLm/E1n$ and $CLh/E1n$ = average quality and top quality. $TCit$ = Google Scholar total citations ; G = Google Scholar G-index.

2.5. The impact of specialisation choices

In this second step, we test the robustness of section 2.4's results and analyse the effect of specialisation choices by including the shares of research output published in each fields (JEL codes at the first letter level) as control variable. This tests whether researchers specialised in some fields publish more or are more cited. [13].

2.5.1. Specialisation only

We first start by regressing publication scores and citation indexes on specialisation shares only. As seen in Table 2.5, French economists specialised in the fields of "Gen-

13. The number of observations in the regressions slightly drops because some researchers have published only articles for which JEL codes are not recorded in EconLit

Table 2.5: Determinants of publication scores and citation indexes: specialisation

	E1n	CLm	CLh	CLm/E1n	CLh/E1n	TCit	G
% A:General	0.736^c	0.418	-0.116	-0.318	-0.852	0.990	0.455
	(0.43)	(0.57)	(1.04)	(0.27)	(0.78)	(0.73)	(0.41)
% B:Thought	0.401^a	-0.127	-1.751^a	-0.528^a	-2.152^a	-0.770^a	-0.537^a
	(0.13)	(0.13)	(0.21)	(0.07)	(0.18)	(0.21)	(0.12)
% C:Maths.	0.104	1.172^a	3.548^a	1.068^a	3.444^a	0.747^c	0.351
	(0.24)	(0.35)	(0.70)	(0.16)	(0.52)	(0.43)	(0.23)
% D:Micro.	0.506^a	1.344^a	3.466^a	0.838^a	2.959^a	0.645^a	0.358^a
	(0.11)	(0.16)	(0.31)	(0.07)	(0.23)	(0.19)	(0.10)
% E:Macro.	0.642^a	0.873^a	1.427^a	0.231^b	0.785^a	-0.028	0.023
	(0.16)	(0.20)	(0.39)	(0.09)	(0.30)	(0.24)	(0.13)
% F:Inter.	0.030	0.039	0.097	0.009	0.067	-0.240	-0.106
	(0.14)	(0.19)	(0.36)	(0.09)	(0.28)	(0.26)	(0.14)
% G:Finance	-0.112	-0.062	-0.084	0.050	0.028	-0.401^b	-0.228^b
	(0.11)	(0.14)	(0.26)	(0.07)	(0.21)	(0.19)	(0.11)
% H:Public	0.070	0.398	1.513^a	0.329^a	1.443^a	-0.027	0.044
	(0.21)	(0.27)	(0.55)	(0.13)	(0.44)	(0.34)	(0.19)
% I:Health	-0.467^b	-0.428^c	-0.450	0.039	0.017	-0.084	-0.034
	(0.21)	(0.26)	(0.47)	(0.10)	(0.35)	(0.32)	(0.20)
% J:Labor	0.146	0.429^a	1.301^a	0.283^a	1.155^a	0.431^b	0.251^b
	(0.12)	(0.16)	(0.33)	(0.08)	(0.27)	(0.20)	(0.12)
% K:Law	-0.304	-0.616	-1.065	-0.312^c	-0.761	-0.722	-0.308
	(0.33)	(0.41)	(0.72)	(0.17)	(0.53)	(0.63)	(0.32)
% L:I.O.	-0.156	-0.275^c	-0.467	-0.119^c	-0.311	0.205	0.080
	(0.12)	(0.16)	(0.31)	(0.07)	(0.23)	(0.21)	(0.12)
% M:Business	-1.059^a	-1.641^a	-3.304^a	-0.582^a	-2.245^a	-0.021	-0.015
	(0.13)	(0.16)	(0.27)	(0.07)	(0.21)	(0.34)	(0.19)
% N:History	-0.513^c	-1.075^b	-2.314^b	-0.562^a	-1.801^a	-1.128^c	-0.693^c
	(0.31)	(0.44)	(0.90)	(0.20)	(0.70)	(0.63)	(0.39)
% O:Develop.	-0.133	-0.245	-0.720^b	-0.112	-0.587^b	-0.005	0.042
	(0.13)	(0.17)	(0.32)	(0.08)	(0.24)	(0.25)	(0.14)
% P:Systems	0.163	-0.134	-1.094^b	-0.297^b	-1.257^a	-0.306	-0.205
	(0.27)	(0.31)	(0.53)	(0.13)	(0.39)	(0.48)	(0.26)
% Q:Agr.	-0.024	0.090	0.595^c	0.114	0.619^a	0.404^b	0.421^a
	(0.12)	(0.16)	(0.31)	(0.07)	(0.24)	(0.19)	(0.11)
% R:Urban	-0.031	-0.161	-0.582^c	-0.130^c	-0.551^b	0.309	0.102
	(0.13)	(0.16)	(0.33)	(0.08)	(0.27)	(0.23)	(0.14)
Constant	0.986^a	3.113^a	-1.423^a	2.127^a	-2.409^a	3.262^a	1.727^a
	(0.04)	(0.05)	(0.10)	(0.02)	(0.08)	(0.08)	(0.04)
R^2	0.058	0.138	0.195	0.085	0.160	0.029	0.034
Observations	1923	1923	1923	1923	1923	1835	1835

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. $E1n$ = number of articles equivalent written alone in EconLit. CLm and CLh are publication scores with a low and high degree of convexity in the journals weighting scheme, respectively. $CLm/E1n$ and $CLh/E1n$ = average quality and top quality. $TCit$ = Google Scholar total citations ; G = Google Scholar G-index.

eral Economics and Teaching" (A), "Macroeconomics and Monetary Economics" (E), "Microeconomics" (D) and "History of Economic Thought, Methodology, and Heterodox Approaches" (B) have published more articles than the average (by order of magnitude of the coefficients). At the other bound, researchers specialised in "Business Administration and Business Economics; Marketing; Accounting" (M), "Economic History" (N) and "Health, Education, and Welfare" (I) have published less articles equivalent written alone.

The French economists average quality of publications is higher in the fields of "Mathematical and Quantitative Methods" (C), "Microeconomics" (D), "Public Economics" (H), "Labor and Demographic Economics" (J) and "Macroeconomics and Monetary Economics" (E) (by order of magnitude of the coefficients). Fields in which total publication scores are higher than the average are the same ones (except for Public Economics (H) in CLm), which means that overrepresented fields in terms of quantity as "General Economics and Teaching" (A) or "History of Economic Thought, Methodology, and Heterodox Approaches" (B) are published in low quality journals.

In terms of Google Scholar citation indexes, French economists specialised in the fields of "Microeconomics" (D), "Labor and Demographic Economics" (J), "Agricultural and Natural Resource Economics; Environmental and Ecological Economics" (Q) and "Mathematical and Quantitative Methods" (C) (only considering total citations for the latter) are more cited than the average. At the other extreme, French economists specialised in the fields of "History of Economic Thought, Methodology, and Heterodox Approaches" (B), "Financial Economics" (G) and "Economic History" (N) are less cited than the average.

Interestingly, comparing Table 2.5 with Table 2.3 in section 2.4, we observe that the R^2 of the model is higher with specialisation shares only than with individual characteristics only to explain total publication scores. Indeed, the total explained variance by specialisation is 13.8% for CLm and 19.5% for CLh whereas the total explained variance by individual characteristics (Table 2.3) is 3.4% for CLm and 6.6% for CLh. By contrast the R^2 of the model explaining citation indexes is higher when we introduce only individual characteristics (8.2% and 10.1% for Google Scholar total citations and G-index, respectively) than when we introduce specialisation choices only (2.9% and 3.4% for total citations and G-index, respectively). In any case, the variance explained by individual characteristics is weak relatively to what is found in the literature on wage equations for instance (in which individual characteristics explain between 30 and 40% of wages, a measure of labour productivity). In academic research, specialisation choices appear to be a stronger determinant of publication outcome than the individual variables we considered in section 2.4 while the reverse is true for citations. We need to assess now whether the conclusions we presented are driven by the omitted specialisation variables or if the two sets of variables corresponds to different effects.

2.5.2. Specialisation and individual characteristics

We introduce now both specialisation patterns and individual characteristics. We conclude from Table 2.6 that the average number of authors per article is the only individual characteristic that matters to explain total publication scores and citation indexes. As found in section 2.4, it has a negative impact on the quantity of published articles and a positive impact on average quality of publications, total publication scores (except in CLm) and citation indexes. Coefficients associated with age, its square and gender are never significant anymore, underlying that demographic impact on total publication scores observed in section 2.4 were in fact due to specialisation choices.

This contrasts with results on the average quality of publications that is still impacted, by the same order of magnitude, by increasing returns to quantity and size of the coauthorship network. Then, we observe that French researchers in economics specialised in the fields of "Microeconomics" (D) and "Labor and Demographic Economics" (J) do publish more articles, of a higher average quality, hence reach higher total publication scores and are more cited. French economists specialised in the fields of "Macroeconomics and Monetary Economics" (E), "Mathematical and Quantitative Methods" (C) and "Public Economics" (H) do also have higher publication scores but are not more cited than the average. At the other bound, researchers specialised in the fields of "History of Economic Thought, Methodology, and Heterodox Approaches" (B) and "Economic History" (N) do publish lower quality articles and are less cited than the average whereas researchers specialised in "Business Administration and Business Economics; Marketing; Accounting" (M) and "Urban, Rural, and Regional Economics" (R) also publish lower quality articles but are not less cited.

Table 2.6: Determinants of publication scores and citation indexes: specialisation and individual characteristics

	E1n	CLm	CLh	CLm/E1n	CLh/E1n	TCit	G
Women	-0.390	-0.714	-2.239	-0.278	-1.597^{c}	0.236	0.094
	(0.62)	(0.84)	(1.68)	(0.29)	(0.89)	(0.84)	(0.46)
Age	0.050	0.039	0.079	-0.022	-0.018	0.051	0.027
	(0.05)	(0.07)	(0.14)	(0.03)	(0.09)	(0.05)	(0.03)
Age^2	-0.000	-0.000	-0.001	0.000	-0.001	0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number	-0.338^{a}	0.094	1.232^{a}	0.203^{b}	0.880^{a}	1.025^{a}	0.954^{a}
	(0.08)	(0.10)	(0.19)	(0.09)	(0.27)	(0.15)	(0.09)
Quantity				0.129^{a}	0.685^{a}		
				(0.03)	(0.08)		
Network size				0.167^{a}	0.565^{a}		
				(0.04)	(0.14)		
% A:General	0.273	0.099	-0.168	-0.189	-0.559	0.400	0.222
	(0.36)	(0.53)	(1.06)	(0.28)	(0.83)	(0.73)	(0.39)
% B:Thought	0.216^{c}	-0.146	-1.355^{a}	-0.374^{a}	-1.665^{a}	-0.631^{a}	-0.346^{a}
	(0.12)	(0.13)	(0.23)	(0.07)	(0.21)	(0.22)	(0.12)
% C:Maths.	0.080	0.978^{a}	2.865^{a}	0.871^{a}	2.675^{a}	0.318	0.043
	(0.22)	(0.34)	(0.69)	(0.14)	(0.41)	(0.39)	(0.20)
% D:Micro.	0.581^{a}	1.354^{a}	3.326^{a}	0.640^{a}	2.149^{a}	0.731^{a}	0.377^{a}
	(0.11)	(0.16)	(0.31)	(0.06)	(0.19)	(0.17)	(0.10)
% E:Macro.	0.602^{a}	0.796^{a}	1.244^{a}	0.081	0.109	-0.016	0.050
	(0.16)	(0.20)	(0.38)	(0.08)	(0.25)	(0.23)	(0.13)
% F:Inter.	0.022	0.059	0.200	0.027	0.137	-0.200	-0.069
	(0.14)	(0.19)	(0.36)	(0.08)	(0.23)	(0.25)	(0.13)
% G:Finance	-0.004	0.040	0.030	0.041	0.025	-0.215	-0.141
	(0.10)	(0.13)	(0.26)	(0.06)	(0.20)	(0.17)	(0.09)
% H:Public	0.106	0.379	1.356^{b}	0.238^{b}	1.106^{a}	-0.006	0.028
	(0.21)	(0.27)	(0.55)	(0.11)	(0.37)	(0.32)	(0.18)
% I:Health	-0.352^{c}	-0.365	-0.513	0.076	0.225	-0.165	-0.142
	(0.19)	(0.25)	(0.49)	(0.09)	(0.30)	(0.30)	(0.18)
% J:Labor	0.332^{a}	0.592^{a}	1.452^{a}	0.172^{b}	0.737^{a}	0.634^{a}	0.309^{a}
	(0.11)	(0.16)	(0.32)	(0.08)	(0.23)	(0.19)	(0.11)
% K:Law	-0.140	-0.427	-0.753	-0.257	-0.475	-0.284	-0.055
	(0.33)	(0.40)	(0.72)	(0.18)	(0.58)	(0.51)	(0.26)
% L:I.O.	-0.083	-0.247	-0.539^{c}	-0.151^{b}	-0.394^{b}	0.204	0.037
	(0.12)	(0.16)	(0.30)	(0.06)	(0.19)	(0.19)	(0.11)
% M:Business	-1.102^{a}	-1.657^{a}	-3.249^{a}	-0.325^{a}	-1.097^{a}	-0.137	-0.111
	(0.12)	(0.16)	(0.30)	(0.08)	(0.22)	(0.31)	(0.18)
% N:History	-0.621^{c}	-1.123^{b}	-2.215^{b}	-0.370^{b}	-0.990^{c}	-1.065^{b}	-0.550^{b}
	(0.35)	(0.47)	(0.94)	(0.18)	(0.58)	(0.48)	(0.26)
% O:Develop.	-0.081	-0.190	-0.643^{b}	-0.092	-0.483^{b}	0.060	0.078
	(0.13)	(0.17)	(0.32)	(0.07)	(0.22)	(0.25)	(0.14)
% P:Systems	0.051	-0.093	-0.679	-0.171	-0.834^{b}	-0.126	-0.022
	(0.27)	(0.32)	(0.55)	(0.12)	(0.37)	(0.49)	(0.25)
% Q:Agr.	0.094	0.073	0.229	-0.059	-0.019	0.171	0.194^{c}
	(0.13)	(0.17)	(0.32)	(0.07)	(0.21)	(0.20)	(0.11)
% R:Urban	0.025	-0.124	-0.588^{c}	-0.157^{b}	-0.646^{a}	0.328	0.100
	(0.13)	(0.16)	(0.32)	(0.07)	(0.24)	(0.22)	(0.14)
Selection	0.374	1.318	6.954	1.088	6.972	-7.935	-3.857
	(3.26)	(4.42)	(8.77)	(1.49)	(4.56)	(9.93)	(5.48)
Constant	-0.454	1.722	-4.994	2.278^{b}	-4.078	0.934	0.190
	(1.74)	(2.36)	(4.69)	(0.90)	(2.70)	(1.51)	(0.84)
R^2	0.143	0.109	0.239	0.457	0.570	0.147	0.176
Observations	1923	1923	1923	1923	1923	1835	1835

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. $E1n$ = number of articles equivalent written alone in EconLit. CLm and CLh are publication scores with a low and high degree of convexity in the journals weighting scheme, respectively. $CLm/E1n$ and $CLh/E1n$ = average quality and top quality. $TCit$ = Google Scholar total citations ; G = Google Scholar G-index. "Selection" is the inverse of Mill's ratio from columns 1 (for publications) and 7 (for citations) of Table 2.4.

2.6. The patterns of knowledge diffusion

Finally, we compare citation indexes and publication scores to see if they measure the same dimensions of research productivity. To do so, we regress Google Scholar total citations indexes and G-indexes on EconLit publication scores, specialisation patterns and individual characteristics including coauthorship variables, which allows us to analyse the patterns of knowledge diffusion.

2.6.1. Citations regressed on publications and specialisation

We start by regressing citation indexes on publication scores (quantity and average quality of publications) and specialisation shares to see if citations' patterns differ from a field to another. Unsurprisingly, researchers who publish more and whose articles are of higher quality are more cited. This confirms that Google Scholar citation indexes capture both the quantity and quality aspects of publication records. In Table 2.7, the total R^2 of the model is slightly higher when the average quality of publications is measured with a high degree of convexity in the journals weighting scheme. With this measure, having published five articles equivalent written alone instead of four increases Google Scholar total citations by 22.4% and Google Scholar G-index by 11.9%.[14] On top of that, increasing the average publication top quality by 10% increases Google Scholar total citations by 2.1% and Google Scholar G-index by 1.2%.[15]

Controlling for publications scores, researchers specialised in the fields of "Business Administration and Business Economics; Marketing; Accounting" (M), "Industrial Orga-

14. $1.25^{0.907} - 1$ and $1.25^{0.503} - 1$, respectively.
15. $1.1^{0.221} - 1$ and $1.1^{0.126} - 1$, respectively.

nization" (L), "Agricultural and Natural Resource Economics; Environmental and Ecological Economics" (Q) and "Urban, Rural, and Regional Economics" (R) (in the order of magnitude of the coefficients) are relatively more cited. At the other bound, researchers specialised in the fields of "History of Economic Thought, Methodology, and Heterodox Approaches" (B), "Macroeconomics and Monetary Economics" (E), "Microeconomics" (D), "Public Economics" (H) and "Financial Economics" (G) are relatively less cited, given their level of publications.

Table 2.7: Determinants of citation indexes: publication scores and specialisation

	TCit	G	TCit	G	TCit	G
Quantity	1.086[a]	0.605[a]	0.970[a]	0.543[a]	0.907[a]	0.503[a]
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Average quality			0.697[a]	0.377[a]		
			(0.05)	(0.03)		
Average top quality					0.221[a]	0.126[a]
					(0.02)	(0.01)
% A:General	0.103	-0.040	0.498	0.174	0.487	0.179
	(0.72)	(0.39)	(0.76)	(0.41)	(0.76)	(0.42)
% B:Thought	-1.156[a]	-0.752[a]	-0.732[a]	-0.523[a]	-0.597[a]	-0.434[a]
	(0.19)	(0.11)	(0.18)	(0.11)	(0.18)	(0.11)
% C:Maths.	0.712[b]	0.332[b]	-0.029	-0.069	-0.032	-0.092
	(0.30)	(0.15)	(0.30)	(0.16)	(0.30)	(0.15)
% D:Micro.	0.101	0.054	-0.464[a]	-0.251[a]	-0.504[a]	-0.291[a]
	(0.14)	(0.08)	(0.15)	(0.08)	(0.15)	(0.08)
% E:Macro.	-0.699[a]	-0.351[a]	-0.780[a]	-0.395[a]	-0.752[a]	-0.381[a]
	(0.20)	(0.11)	(0.19)	(0.10)	(0.19)	(0.11)
% F:Inter.	-0.265	-0.119	-0.276	-0.126	-0.293	-0.136
	(0.21)	(0.12)	(0.22)	(0.12)	(0.21)	(0.11)
% G:Finance	-0.225	-0.130	-0.273[c]	-0.156[c]	-0.250[c]	-0.144[c]
	(0.16)	(0.09)	(0.15)	(0.08)	(0.15)	(0.08)
% H:Public	-0.223	-0.065	-0.482[c]	-0.205	-0.570[b]	-0.263[c]
	(0.28)	(0.16)	(0.27)	(0.15)	(0.27)	(0.15)
% I:Health	0.332	0.198	0.283	0.171	0.272	0.163
	(0.22)	(0.13)	(0.23)	(0.14)	(0.22)	(0.14)
% J:Labor	0.349[b]	0.205[b]	0.178	0.112	0.130	0.080
	(0.15)	(0.09)	(0.15)	(0.09)	(0.15)	(0.09)
% K:Law	-0.405	-0.132	-0.238	-0.041	-0.313	-0.079
	(0.56)	(0.29)	(0.53)	(0.28)	(0.51)	(0.27)
% L:I.O.	0.418[b]	0.198[c]	0.493[a]	0.239[b]	0.467[a]	0.226[b]
	(0.18)	(0.10)	(0.18)	(0.10)	(0.18)	(0.10)
% M:Business	1.203[a]	0.667[a]	1.504[a]	0.831[a]	1.525[a]	0.851[a]
	(0.34)	(0.19)	(0.34)	(0.19)	(0.33)	(0.19)
% N:History	-0.854	-0.540	-0.576	-0.389	-0.570	-0.378
	(0.60)	(0.39)	(0.57)	(0.37)	(0.58)	(0.37)
% O:Develop.	0.166	0.138	0.227	0.170	0.278	0.201[c]
	(0.21)	(0.12)	(0.21)	(0.12)	(0.20)	(0.11)
% P:Systems	-0.393	-0.254	-0.172	-0.134	-0.087	-0.079
	(0.31)	(0.17)	(0.31)	(0.17)	(0.30)	(0.16)
% Q:Agr.	0.422[b]	0.431[a]	0.329[b]	0.380[a]	0.273	0.346[a]
	(0.17)	(0.09)	(0.17)	(0.09)	(0.17)	(0.09)
% R:Urban	0.413[b]	0.160	0.509[b]	0.212[c]	0.536[a]	0.230[c]
	(0.21)	(0.13)	(0.20)	(0.12)	(0.20)	(0.12)
Constant	2.126[a]	1.094[a]	0.752[a]	0.349[a]	2.827[a]	1.494[a]
	(0.08)	(0.04)	(0.14)	(0.08)	(0.09)	(0.05)
R^2	0.426	0.423	0.481	0.474	0.482	0.481
Observations	1835	1835	1835	1835	1835	1835

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. Quantity = $E1n$, number of articles equivalent written alone in EconLit. Average quality and top quality = $CLm/E1n$ and $CLh/E1n$. $TCit$ = Google Scholar total citations ; G = Google Scholar G-index.

2.6.2. Citations regressed on publications, specialisation and individual characteristics

Finally, we regress citation indexes on publication scores (quantity and average quality of publications), specialisation and individual characteristics including coauthorship variables. Results of section 2.6.1 are robust to the introduction of individual characteristics except that researchers specialised in “Agricultural and Natural Resource Economics; Environmental and Ecological Economics” (Q) are not over-cited and researchers specialised in “Financial Economics” (G) are not under-cited anymore. Controlling for quantity of published papers and average quality of these publications, older researchers have slightly higher total Google Scholar citations score, which could be due to the fact that their articles were published longer ago and hence have had more time to be cited. But they do not have higher G-indexes. Importantly, at given publication records, researchers with higher average number of authors per article and with larger network size (total number of different coauthors) are significantly more cited. Our interpretation of this result is that larger team sizes generate more knowledge diffusion through conferences, seminars, informal talks, etc. Also, researchers might generate knowledge diffusion of their new publications through their former coauthors. With average quality (and not top quality) as a determinant of citation indexes, a publisher that has on average two coauthors instead of one only (per article) has been cited 9.7% more and has a G-index 20.5% higher.[16] Moreover, having a stock of five different coauthors instead of four, meaning a total network size of 6 instead of 5, increases Google Scholar total citations by 5.7% and Google Scholar G-index by 4.2%.[17]

16. $1.5^{0.229} - 1$ and $1.5^{0.460} - 1$, respectively.
17. $1.2^{0.303} - 1$ and $1.2^{0.225} - 1$, respectively.

We perform a variance analysis to infer the relative explanatory power of each variable (or group of variables). In order to do so, we first calculate the effect of each variable as the product of the variable and its coefficient. Then, we calculate the standard error of this effect on all observations and the correlation coefficients between the calculated effects and the dependent variable. To have an important explanatory power of the dependent variable, the effect of an explicative variable should first have a high standard error by comparison with the standard error of the dependent variable. However, and most importantly, a variable or a group of variable, has a large explanatory power when its effect is largely correlated with the dependent variable.

In Table 2.9, we perform the variance analysis of the estimations reported in columns 3 and 5 of Table 2.8, where the log of Google Scholar total citations is the dependent variable. In Table 2.10, we perform the variance analysis of columns 4 and 6 of Table 2.8, where the log of Google Scholar G-index is the dependent variable. In both cases, with a standard error equal to around $55 - 60\%$ of the dependent variable and a correlation coefficient of around 0.65 with it, EconLit publication scores are the most important determinants of citation indexes. [18] In this group of variables, the quantity of publications has a much higher explanatory power than the average quality of these publications with a standard error between 1.5 times and twice larger and a correlation coefficient 1.5 times larger, whatever the weighting scheme of journals. Then, we observe that other explicative variables have much weaker explanatory powers of citation indexes than publications scores. For total Google Scholar citations, even if individual demographic characteristics (gender and age) have a higher standard error than coauthorship effects, the correlation coefficient of the

18. $\frac{1.021}{1.756} = 0.581$, $\frac{1.037}{1.756} = 0.591$, $\frac{0.532}{0.988} = 0.538$, $\frac{0.544}{0.988} = 0.551$

latter with the dependent variables are 1.8 times larger than the correlation coefficients of individual demographic characteristics. [19] For Google Scholar G-indexes, it is even clearer that coauthorship patterns matter more than individual demographic characteristics since both standard errors and correlations coefficients with the dependent variables are higher for coauthorship effects than for individual demographic characteristics (importantly, correlation coefficients are more than twice larger). [20] Hence, the determinants of publication productivity seems to differ from those in market activities and encompass network effects that are typical of this activity.

In the individual demographic characteristics group of variables, age has the largest explanatory power. In the coauthorship patterns group of variables, network size has a larger explanatory power than the average number of authors per article. Knowledge diffusion is driven more by total coauthorskip network than by team sizes. Finally, even if we observe some fields disparities in the citation practices described above, we observe in Tables 2.9 and 2.10 that the measured effects of specialisation shares are very weakly correlated with the dependent variables, meaning that specialisation patterns have a weak explanatory power of citation indexes like Google Scholar total citations and G-indexes once all individual effects are controlled for. In both Tables, using CLm or CLh to measure journals quality leads to very similar results.

19. $\frac{0.514}{0.284} = 1.810$, $\frac{0.516}{0.283} = 1.823$
20. $\frac{0.539}{0.256} = 2.105$, $\frac{0.537}{0.255} = 2.106$

Table 2.8: Determinants of citation indexes: publication scores, specialisation and individual characteristics

	TCit	G	TCit	G	TCit	G
Quantity	0.804[a]	0.434[a]	0.744[a]	0.407[a]	0.693[a]	0.381[a]
	(0.05)	(0.03)	(0.04)	(0.02)	(0.05)	(0.02)
Average quality			0.681[a]	0.314[a]		
			(0.05)	(0.03)		
Average top quality					0.216[a]	0.104[a]
					(0.02)	(0.01)
Women	-0.119[c]	-0.052	-0.070	-0.029	-0.063	-0.025
	(0.07)	(0.04)	(0.07)	(0.04)	(0.07)	(0.04)
Age	0.019	0.007	0.048[c]	0.020	0.050[c]	0.022
	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)	(0.01)
Age^2	0.000	0.000	-0.000	-0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number	0.342[b]	0.512[a]	0.229	0.460[a]	0.199	0.443[a]
	(0.16)	(0.09)	(0.16)	(0.09)	(0.16)	(0.09)
Network size	0.452[a]	0.294[a]	0.303[a]	0.225[a]	0.280[a]	0.211[a]
	(0.06)	(0.04)	(0.06)	(0.03)	(0.06)	(0.03)
% A:General	0.213	0.135	0.404	0.223	0.368	0.210
	(0.68)	(0.35)	(0.74)	(0.37)	(0.74)	(0.38)
% B:Thought	-0.747[a]	-0.396[a]	-0.503[a]	-0.283[a]	-0.399[b]	-0.229[b]
	(0.19)	(0.10)	(0.19)	(0.10)	(0.19)	(0.10)
% C:Maths.	0.274	0.006	-0.338	-0.277[c]	-0.320	-0.280[c]
	(0.28)	(0.14)	(0.29)	(0.15)	(0.28)	(0.14)
% D:Micro.	0.096	0.014	-0.376[a]	-0.204[a]	-0.406[a]	-0.227[a]
	(0.14)	(0.07)	(0.14)	(0.07)	(0.14)	(0.08)
% E:Macro.	-0.610[a]	-0.284[a]	-0.677[a]	-0.315[a]	-0.650[a]	-0.304[a]
	(0.19)	(0.11)	(0.18)	(0.10)	(0.18)	(0.10)
% F:Inter.	-0.267	-0.112	-0.285	-0.120	-0.302	-0.129
	(0.20)	(0.10)	(0.21)	(0.10)	(0.20)	(0.10)
% G:Finance	-0.171	-0.116	-0.200	-0.129[c]	-0.177	-0.118
	(0.14)	(0.07)	(0.13)	(0.07)	(0.13)	(0.07)
% H:Public	-0.293	-0.138	-0.475[c]	-0.223	-0.552[b]	-0.263[c]
	(0.27)	(0.15)	(0.25)	(0.14)	(0.25)	(0.14)
% I:Health	0.200	0.066	0.177	0.056	0.173	0.053
	(0.23)	(0.14)	(0.22)	(0.14)	(0.22)	(0.14)
% J:Labor	0.362[b]	0.155[c]	0.261[c]	0.108	0.224	0.088
	(0.15)	(0.08)	(0.14)	(0.08)	(0.14)	(0.08)
% K:Law	-0.161	0.016	0.004	0.092	-0.073	0.058
	(0.50)	(0.25)	(0.46)	(0.23)	(0.44)	(0.23)
% L:I.O.	0.342[b]	0.114	0.462[a]	0.169[c]	0.444[a]	0.163[c]
	(0.17)	(0.09)	(0.17)	(0.10)	(0.17)	(0.10)
% M:Business	1.189[a]	0.649[a]	1.451[a]	0.770[a]	1.469[a]	0.783[a]
	(0.32)	(0.18)	(0.32)	(0.18)	(0.31)	(0.18)
% N:History	-0.759[c]	-0.385	-0.563	-0.294	-0.571	-0.294
	(0.41)	(0.26)	(0.41)	(0.25)	(0.42)	(0.26)
% O:Develop.	0.196	0.155	0.250	0.179[c]	0.299	0.204[c]
	(0.20)	(0.11)	(0.20)	(0.11)	(0.20)	(0.11)
% P:Systems	-0.252	-0.094	-0.135	-0.040	-0.069	-0.006
	(0.31)	(0.16)	(0.32)	(0.16)	(0.31)	(0.16)
% Q:Agr.	0.035	0.105	0.077	0.124	0.047	0.110
	(0.17)	(0.09)	(0.16)	(0.09)	(0.16)	(0.09)
% R:Urban	0.353[c]	0.111	0.466[b]	0.163	0.495[b]	0.179
	(0.20)	(0.12)	(0.19)	(0.11)	(0.19)	(0.11)
Constant	0.670	0.153	-1.316[b]	-0.765[b]	0.682	0.159
	(0.65)	(0.35)	(0.61)	(0.34)	(0.59)	(0.32)
R^2	0.482	0.517	0.528	0.548	0.527	0.550
Observations	1835	1835	1835	1835	1835	1835

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. Quantity = $E1n$, number of articles equivalent written alone in EconLit. Average quality and top quality = $CLm/E1n$ and $CLh/E1n$. $TCit$ = Google Scholar total citations ; G = Google Scholar G-index.

Table 2.9: Variance analysis, Google Scholar total citations

	CLm as quality		*CLh* as top quality	
	Stand. err.	Correlation	Stand. err.	Correlation
Explained: GS Total citations	1.756	1.000	1.756	1.000
EconLit Publications	1.021	0.659	1.037	0.662
Quantity	0.782	0.618	0.728	0.618
Average quality	0.465	0.410	.	.
Average top quality	.	.	0.501	0.472
Individual effects	0.328	0.284	0.332	0.283
Gender	0.031	0.171	0.028	0.171
Age	0.320	0.275	0.325	0.275
Coauthorship	0.287	0.514	0.262	0.516
Authors number	0.062	0.157	0.054	0.157
Network size	0.251	0.547	0.232	0.547
Specialisation	0.333	0.040	0.330	0.031
Residuals	1.206	0.687	1.207	0.688

Table 2.10: Variance analysis, Google Scholar G-index

	CLm as quality		*CLh* as top quality	
	Stand. err.	Correlation	Stand. err.	Correlation
Explained: G-index	0.988	1.000	0.988	1.000
EconLit Publications	0.532	0.653	0.544	0.660
Quantity	0.427	0.611	0.400	0.611
Average quality	0.215	0.404	.	.
Average top quality	.	.	0.241	0.476
Individual effects	0.160	0.256	0.163	0.255
Gender	0.013	0.159	0.011	0.159
Age	0.157	0.249	0.160	0.249
Coauthorship	0.270	0.539	0.256	0.537
Authors number	0.125	0.258	0.120	0.258
Network size	0.187	0.607	0.175	0.607
Specialisation	0.169	0.043	0.169	0.031
Residuals	0.664	0.672	0.663	0.671

2.7. Conclusion

We study the individual determinants of EconLit publication scores and Google Scholar citation indexes of French academic economists. We show that once controlled by coauthorship patterns (average number of authors per article and total coauthors' network size), demographic characteristics (gender and age) do not matter anymore, except for the probabilities to publish and to be cited at least once. Moreover, we have carefully analysed specialisation patterns of French academic economists and researchers specialised in the fields of Microeconomics and Labor and Demographic Economics do publish more articles, of a higher average quality, hence reach higher total publication scores and are more cited.

Importantly, we exhibit increasing returns to quantity and coauthors' network size for the average quality of publications, whatever the chosen weighting scheme of journals, and increasing returns to the average number of authors per article for all research output taking quality into account, including citation indexes. Finally, by looking at the pattern of knowledge diffusion, we find that publication scores are the most important determinant of citation indexes followed by coauthorship characteristics. Researchers who publish more papers and of higher quality are more cited, which was expected. Now, we also show that it is also the case for researchers working in larger coauthor's teams and who have a large total coauthorship network. Our interpretation of these results is the following. Coauthored articles are presented in conferences, seminars and workshops by their several authors. Moreover researchers discuss their new findings with their peers, including those with whom they have already worked. Both attitudes generate more knowledge diffusion that we can measure through citations. Therefore confirms the intuition usually widely spread that network matters for citations.

Hence, given both that citations have been proved to be an important determinant of academic wages and that network effects matter for academic promotions (see McDowell and Smith (1992), Combes, Linnemer and Visser (2008) or Zinovyevay and Bagues (2012) for instance) and citations (as seen here), having wage data would allow to disentangle between direct network effects on wages and indirect effects through citations. Also, looking at the sources of citations (who are the citing researchers/articles?) would allow to have a more precise picture of the patterns of knowledge diffusion. Finally, it would be interesting to perform some structural estimations or to benefit from richer datasets (possibly panel, with workable instruments for endogeneity concerns) to infer directions of causalities, as for instance Bramoullé, Djebbari and Fortin (2009) propose to identify peer effects in recreational services. This would improve on the correlations established here and move us to more causal interpretations. This is a difficult exercise though, which is left for further research.

Part II

Two Essays in Economics of Science: Economic Geography and Labour Economics

Chapter 3

Do Large Departments make Academics more Productive? Agglomeration and Peer Effects in Research [1]

We study the role on individual publication records of a large set of departments' characteristics controlling for many individual time-varying characteristics, individual fixed-

1. This chapter has been co-written with Pierre-Philippe Combes. It has benefited from the excellent research assistance of Philippe Donnay and Charles Laitong to whom we are very grateful. We also thank Yann Bramoullé, Gilles Duranton, Laurent Gobillon, Jean-Marc Robin, Maurizio Iacopetta, Frédéric Robert-Nicoud, Elisabet Viladecans-Marsal, Cecilia García-Peñalosa, Alain Trannoy as well as participants to the First Workshop on Urban Economics (Barcelona University), the 25th EEA Annual Congress (Amsterdam), the 59th AFSE Annual Congress (Paris), the 57th NARSC Annual Meetings (Denver), the 11th RIEF Doctoral Meeting (Nice) and the 10th Journées LAGV (Marseille) for useful suggestions. The Aix-Marseille School of Economics is a new "Laboratoire d'Excellence' of Aix-Marseille University (Amu), Centre National de la Recherche Scientifique (Cnrs), Ecole Centrale de Marseille (Ecm), Ecole des Hautes Etudes en Sciences Sociales (Ehess), Institut de Recherche pour le Développement (Ird), and Institut National de la Santé et de la Recherche Médicale (Inserm). This study was made possible and financed by the French Ministry of Research.

effects and reverse causality. Departments' characteristics have an explanatory power that can be as high as individual characteristics' one. Departments generating most externalities are those where academics are homogeneous in terms of publication performance and have diverse research fields, and, to a lower extent, those that are large, with more women, older academics, stars and co-authors connection to foreign departments. More students per academic does not penalise publication. At the individual level, women and older academics publish less, while the average publication quality increases with average number of authors per paper, individual field diversity, number of published papers and connections to foreign departments.

3.1. Introduction

Any academic has an opinion about what makes a good department. Surprisingly enough, there are little hard studies quantifying this precisely although possible implications for an optimal design of education and research policies are numerous. We focus here on the role on individual publication records of both individual characteristics and a large set of department characteristics. We develop a careful strategy that controls for possible spatial selection of academics, reverse causality and missing variables. Clearly, both our identification strategy and the results we get for academic research are relevant more generally, for instance for all knowledge intensive industries where individual abilities play a crucial role (as in R&D departments of many manufacturing industries, or in finance, law or health services).

We propose to answer three sets of questions. First, and at the most general level, what makes an individual productive, her own abilities or the firm in which she works? Applied to science, these translates into whether academics publish more because they have better abilities (gender, age, type of position or any other individual characteristics possibly unobserved) and a more rewarding publication strategy (research field, number and location of co-authors) or because they are located in departments that provide a better local environment with stronger externalities? Using an exhaustive panel of French academics and departments in economics over 19 years (1990-2008) and their quality-adjusted publication records in EconLit, we find that both types of explanations are relevant. In particular, location is an important determinant of the individual quantity and quality of publications and represents at least half of the explanatory power of individual characteristics. The individual strategy that consists in publishing in different fields of economics, with co-authors located abroad, and with a high number of co-authors per paper are the most rewarding in terms of publication quality. The latter result suggests the presence of increasing returns to scale at the co-author team level. The average quality of the publications of an academic also increases with her number of publications, suggesting the presence of increasing returns to scale also at the individual level. Everything else equal (including the field of specialisation and all the aforementioned variables), women and older academics (for a given type of position) publish less.

Then, we move to the dual question that assesses the extent to what, again at the most general level, more productive firms simply attract more productive employees or generate more productive environments. That is, for academia, whether good departments, defined as those where the average quantity and quality of publications per academic is high, are those where highly productive academics locate or those that generate more externalities? Even if we exhibit the presence of some spatial sorting of academics, the most productive academics being located in the departments that generate more externalities, we find that local externalities and composition of departments with more or less productive academics equally matter when explaining the department hierarchy in terms of publications.

Finally, and most importantly for the optimal design of firms or institutions, what are the channels of local external effects? We enter the black box of department externalities and assess the relative magnitude of the channels through which they operate. Having fairly homogenous academics in terms of publication records and a pretty diversified set of research fields (within economics) are the two characteristics that best explain department externalities. Then, the presence of stars (researchers in the top 1% of French academics) and department size, in terms of number of academics, constitute the second set of factors explaining department externalities, as, although to a lower extent, the importance of co-authorship with academics abroad and US academics. Conversely, the proximity, in terms of physical distance, to other French departments, has little impact. All these effects are present as regards both quantity and quality of publications. Importantly, they are stronger for the latter. We also observe significant positive externalities from the share of women or of older academics in the department even if these are not large determinants of the disparities between departments. We do not find any large reverse causality bias when estimations are instrumented (and are performed, due to instrument availability, at a slightly more aggregated geographical scale, the employment area). When the specification is extended to encompass the role of teaching load, which requires to reduce the time span due to data availability, we obtain, possibly against some beliefs, that more students per academic in the department does not affect quantity and quality of publications.

The paper first relates to the new surge of interest over the last ten years in the estimation of agglomeration economies (see for reviews Rosenthal and Strange, 2004, Melo, Graham, and Noland, 2009, Moretti, 2011). Assessing how much is gained by further concentrating economic activities over space, by increasing regional specialisation or concentrating some types of occupations, is indeed a crucial preliminary step to evaluate regional policies. A parallel can be made with universities or government providing incentives to make academic departments larger or more specialised for instance. Simultaneously, another strand of literature recently focus on the role of peer and network effects on various schooling or labour market outcomes (see for reviews Jackson, 2011, Sacerdote, 2011). This is now the role of composition, by gender, ethnic or social origin, of a group, or network, on individual outcomes of its members that is under focus. Clearly, similar questions for departments arise as regards their optimal share of assistant professors over full professors for instance, of women, of elder academics, or of some academics particularly talented or connected to co-authors located in other institutions. For both agglomeration and peer effects, the access to new data sets, typically encompassing information at the individual level (firms, workers, students/pupils or academics as here) and the search for relevant econometric strategies have much enlarged identification possibilities and have clarified the direction of causalities, a trend we also follow here.

A couple of recent papers to which we compare our results consider a sub-set of the effects we identify here. Waldinger (2012) conclude to an absence of localised peer effects in Germany among physicists, mathematicians and chemists at the Nazi time. Kim, Morse and Zingales (2009) reach a similar conclusion that being affiliated to one of the top-25 US university has no more effect on the individual outcomes of academics in economics and finance in the 1990s, while it was the case in the 1970s and 1980s. This is confirmed by

Dubois, Rochet and Schlenker (2011) on mathematicians who show that the best departments do not necessarily stimulate positive externalities even if they are the most successful in hiring the most promising academics. Oyer (2006) show that a top placement for freshly economics PhD has long-term benefits in terms of career but no benefit in terms of enhanced productivity, supporting also the view that top departments have no productivity spillovers (in the 1990s). Therefore, our conclusion that departments explain a large share of academics' individual productivity is somewhat discordant. It can be explained either by the different context under study, which would mean that European institutions at contemporaneous time generate more local externalities than US departments or German universities at the Nazi time, or by the fact that our data set allows us to consider more effects and to develop a more complete econometric analysis.

We decompose individual productivity into three components, the probability to publish in a given period and field, the number of publications and the average quality of these publications and we study the determinants of these three components separately. The literature is more used to considered only a number of publications adjusted for quality as dependent variable. We show that the role of some variables do differ according to one or the other of these productivity measures, which means that the optimal strategy for an individual or a department should depend on the targeted dimension of publications. Some departments may be constrained in terms of characteristics they can directly influence. Therefore, some may choose preferably to increase the share of publishers in the department when others could on the contrary think about specialising more tasks, *i.e.* allocating some people fully to teaching and administration while trying to improve the quantity and quality of papers published by the other ones. Also, given their characteristics in terms of size and specialisation, some departments could find it more efficient to target the quantity of publications more than their quality, or reversely. All of this underlines the importance of characterising the most efficient local structure to improve departments' research productivity on each dimension separately. Moreover, we also perform our estimations on two different indexes of publications quality, taking it slightly or strongly into account. Typically, the (individual and department) determinants of publications in top journals might differ from those in field journals. Our use of all EconLit publications and a corresponding impact factor index for all the 1200 EconLit journals allows us to study such differences when the literature usually restricts itself to a small number of journals (23 journals in Waldinger (2012), 41 in Kim et al. (2009), 68 in Dubois et al. (2011)).

We take very seriously the concern of a possible spatial sorting of talents that would influence the measure of department effects. Such possible selection effects are mentioned in most papers on peer effects (Sacerdote, 2011) and are also central in recent assessments of agglomeration economies in market activities as emphasised by Combes, Duranton and Gobillon (2008). We use individual data to tackle them properly. We run estimations on individual productivity considering both individual and department variables, and both department and individual fixed effects. We cannot use a natural experiment to remove endogenous selection to departments as Waldinger (2012) (who uses the dismissal of scientists in Nazi Germany) or Azoulay, Graff Zivin and Wang (2010) (who uses the premature death of superstars academics) do to identify peer effects. However we provide estimates of agglomeration economies net of the possible academic spatial sorting, would it be based on either observed or unobserved individual and department characteristics, which corresponds to a pretty general model. Moreover, and contrary to the recent papers on peer effects in academia, we also propose some instrumental variables estimations that take into account possible reverse causality issues for some department characteristics. Not using a natural experiment also presents the advantage of providing more general results. For instance, the co-authors of superstars, or the Nazi dismissed scientists, may not share the same characteristics as the average academic. Our data set is not only exhaustive on all academics present in France but it also presents the second important advantage of reporting non-publishing academics. Studies that only use bibliometric sources necessarily ignore them. This means that their department characteristics, computed on publishers only, are affected by some measurement error. For instance, department size is not the number of academics but the number of academics who publish over the period (which is in general short) in the department. Last, we also have more individual characteristics as age, gender or the position hold, which can impact publication output and are usually not present in the data sets used in other studies. All of this can impact the obtained results and explain our new findings.

The rest of the paper is organised as follows. Section 3.2 presents the theoretical framework, the studied variables and the econometric strategy, while data is detailed in section 3.3. Results are presented in sections 3.4, 3.5, and 3.6 for the determinants of individual publications, the variance analysis of departments' performance, and the determinants of departments' performance, respectively. Section 3.7 presents some robustness checks and section 3.8 concludes.

3.2. Theory and estimation

3.2.1. Publication output, individual and local effects

Let y_{ift} denote the publication output adjusted for quality (presented in section 3.3.1) of academic i in field f at date t. The total output of academic i at date t is the sum of y_{ift} over all her fields, $y_{it} = \sum_f y_{ift}$. Some academics sharing their time between many departments, let α_{idt} denote the share of academic i's output attributed to department d at date t.[2] Department d output in field f at date t is given by $Y_{dft} = \sum_{i \in dt} \alpha_{idt}\, y_{ift}$ and its total output by $Y_{dt} = \sum_f Y_{dft} = \sum_{i \in dt} \alpha_{idt}\, y_{it}$.

We assume that y_{ift} is given by

$$y_{ift} = e_{it}\, A_{d(i,t)ft} \ ,$$

where e_{it} is the academic own efficiency at date t and $A_{d(i,t)ft}$ the efficiency in field f of the department academic i belongs to at date t, $d(i,t)$. For academics belonging to many departments, one such specification is assumed for each department, A_{dft} being specific to each one of them in this case.

The observable part of department d efficiency in field f at date t, A_{dft}, is then supposed to depend on two components, as follows:

$$A_{dft} = \text{Composition}_{dt}\ \text{Department Research Strategy}_{dft} \ .$$

Composition$_{dt}$ corresponds to a vector of effects from the demographic structure of the

2. 90% of our academics have only one affiliation in which case these shares α_{idt} are equal to 1.

department. This includes the logarithm of the department size ($\text{Size}_{dt} = \log\left(\sum_{i \in dt} \alpha_{idt}\right)$ the department full-time equivalent number of academics), the average age of academics, the share of women among academics, and the share of the various types of positions in the department.

Department size plays the role of the total employment density variable considered in standard estimations of agglomeration economies that reflect possible local externalities from the overall size of the local economy. The list of possible positive effects from department size is long. As examples among many, academics in larger departments can benefit from larger administrative or research assistance staff, from an extended bargaining power within the university, or at the national level, that allows them to get more research funds, or from a better overall visibility that makes network effects stronger, even if some of these effects are captured by some of the Department Research Strategy$_{dft}$ variables. One cannot exclude the presence of congestion effects that would lead to a negative impact of size.

For given size, departments have younger or older academics, more or less women, or more full professors versus assistant professors for instance. As suggested by Hellerstein, Neumark and Troske (1999), these composition effects have to be introduced in the specification as their shares among the department total number of academics. We assess whether local externalities are stronger from different types of academics. For instance, older academics can make the other academics benefit from their experience, women can generate more externalities than men, and similarly for the various types of positions. A literature in industrial organisation (see for instance Besley and Ghatak, 2008, Auriol, Friebel, and Lammers, 2012) studies the role of status incentives with implications for the optimal share of the different positions within the firms, which is another interpretation of such variables. The French academic system is rather complex in terms of possible academic status. On top on the distinction between lower and upper positions (assistant professor versus full professor), some academics have research obligations while others not, some are attached to the local university while others depend on national research institutes (Cnrs, Inra, Ehess,...). Each type can generate more or less externalities since both time devoted to research and incentives to locally cooperate differ. This leads us to consider 13 different positions in the specification, which are detailed in Appendix 3.D. Moreover most academics of our data set are categorised as pure economists but some of them are also attached to other domains (business, mathematics,...), an information we also have. Hence, we also introduce in the specification the department share of non pure economists, and we do that separately for the university and the Cnrs academics since the definition of an economist is not exactly the same.

The variables considered in the Research Strategy$_{dft}$ vector also allow us to identify different sources of department externalities. First, we evaluate through a specialisation variable, the share of department d's output in field f at date t, the role of what is called 'Localisation economies' by economic geography:

$$\text{Specialisation}_{dft} = \log \frac{Y_{dft}}{Y_{dt}} \ .$$

Marshall (1890) first developed the idea that the relative size of an industry within the local economy can generate stronger local externalities for this industry, for instance when it uses specific local public goods, specific inputs or labour types. The same intuition can be developed for a field in academic research, for instance because all fields within economics are not internationalised to the same extent, or because they do not need the same research mix in terms of research assistance, computer capacities, or access to data. Benefiting from a measure of publication at the field level allows us to test whether academics in departments more specialised in a field publish more in this field.

Conversely, it is argued since Jacobs (1969) that the overall diversity of the local activity can be beneficial to local productivity, in particular in research intensive sectors. There would be some cross-fertilisation of ideas among industries, which would strengthen innovation and growth. A large literature has attempted to test this idea by introducing a diversity index in the estimated specifications, typically a Herfindhal index on the shares of each industry in the local economy. We proceed similarly here with the shares of each Jel code among the department publications. A problem arises because, by construction, such a gross diversity index is highly correlated with department size. This is because departments with few academics have lots of Jel codes without any publication. To remove this size effect absent from standard economic geography studies since there are few locations without any activity in an industry, we subtract from the gross diversity index the value it would take if all academics within the department would have chosen their Jel codes randomly. The diversity index net of size effect is obtained as:

$$\text{Department Diversity}_{dt} = \log \left[\sum_f \left(\frac{Y_{dft}}{Y_{dt}} \right)^2 \right]^{-1} - \log \left[\sum_f \left(\frac{\tilde{Y}_{dft}}{\tilde{Y}_{dt}} \right)^2 \right]^{-1},$$

where $\sum_f \left(\frac{\tilde{Y_{dft}}}{\tilde{Y_{dt}}} \right)^2$ is the randomly generated Herfindhal index built by simulations. [3]

3. We first attribute random Jel codes to each publication assuming that the probability to publish in each Jel code follows a Binomial law with a probability of success given by the share of output in each Jel

The third research strategy effect regards the physical proximity to other departments with which academics could interact or on the contrary compete. We capture this type of effects by an external research access variable, Research Access$_{dt}$. It tell us whether externalities also emerge between different but nearby departments, as it has been emphasised by economic geography for market activities over the last decade. Research Access$_{dt}$ is the spatially discounted sum of research outputs of all other departments:

$$\text{Research Access}_{dt} = \log \sum_{d' \neq d} \frac{Y_{d't}}{Dist_{dd'}} ,$$

where $Dist_{dd'}$ is the geographical distance between departments d and d'.[4]

Departments also differ in terms of co-authorship patterns of their academics. Having academics connected to foreign academic institutions can generate positive externalities through network effects for instance, which has been emphasised recently both in market activities and for research (see Ductor, Fafchamps, Goyal and van der Leij (2011) for a recent example in economics). We compute the share of the department academics connected to (at least one) co-author who is located outside France but not in the USA (Non-USA openness$_{dt}$) and the same share for the department academics connected to co-authors located in the USA (USA openness$_{dt}$).

There are debates in departments about whether hiring top academics is a good strategy for other academics or not. We test more generally the possible role that department heterogeneity in terms of academics publication records plays a positive or a negative role

code at the national level. Then, the department diversity index is recomputed using these new Jel codes. The department randomly generated Herfindhal index is the average of 1,000 such procedures.

4. Alternative specifications of the research access variable, with squared distance or square root of distance in the denominator have been tested and lead to qualitatively similar results as discussed in section 3.6.

on individual publication records. We complement this effect with a specific assessment of the role of stars, in the spirit of Azoulay et al. (2010). Department heterogeneity is measured by the within department coefficient of variation of individual output:

$$\text{Heterogeneity}_{dt} = \log \frac{\text{Standard Deviation}(y_{it})}{\text{Average}(y_{it})} ,$$

where Standard Deviation(y_{it}) and Average(y_{it}) are the standard deviation and the average of individual publication outputs within department d at date t. We also introduce in the specification the department share of academics who are in the top centile of the most productive academics in France, Stars_{dt}.

We move now to the description of individual variables. The observable part of the individual efficiency of academic i at date t, e_{it} is also supposed to depend on two components, as follows:

$$e_{it} = \text{Individual Characteristics}_{it}\ \text{Individual Research Strategy}_{it} .$$

Importantly, the data set we use allows us to identify simultaneously the impact of individual characteristics, and therefore to control for the possible non random selection of academics across departments, and the externality impact of these very characteristics. For instance, older academics might publish less individually while exerting a positive externality on the other academics of the department. Therefore Individual Characteristics$_{it}$ is the vector, at the individual level, of all the variables for which a possible externality at the department level is tested. This includes academic i age and its square, gender, position hold and dummy variables for being connected to (at least one) co-author abroad but not in the USA (Non-USA openness$_{it}$), in the USA (USA openness$_{it}$) and for being a star (here, to be ranked among the French top 1% of academics).

We also include a vector of variables that can be considered as reflecting individual research strategy. To test for the presence of economies of scale within co-author teams, we introduce the average number of authors per article written by academic i at date t, Authors Number$_{it}$. This variable is central in many studies on the determinants of publication records that do not take into account the role of location but evaluates the returns to co-authorship following Sauer (1988). We also consider academic i field diversity, Individual Diversity$_{it}$, to assess whether academics benefit from knowledge acquired in other fields to publish in field f. This tests the presence of complementarities between fields at the individual level:

$$\text{Individual Diversity}_{it} = \log\left[\sum_{f} \mathbb{1}(y_{ift} > 0)\right],$$

where $\mathbb{1}(y_{ift})$ is a dummy variable equal to 1 when academic i production in field f is non zero at date t.

3.2.2. Econometric specifications

We follow the econometric strategy to separate agglomeration and peer effects from individual characteristics proposed by Combes, Duranton and Gobillon (2008). It is a two-step procedure in which, in the first step, the logarithm of individual productivity in a given field (y_{ift}) is regressed on individual effects (and possibly an individual fixed effect), a department-time fixed effect (β_{dt}) and the department research strategy variables that depend on the field, which reduces here to $\text{Specialisation}_{dft}$:

$$\begin{aligned} \log\ y_{ift} \ &= \ \theta_i \ + \ \text{Individual Characteristics}_{it}\varphi \ + \ \text{Individual Research Strategy}_{it}\phi \ + \\ & \beta_{d(i,t)t} \ + \ \text{Specialisation}_{d(i,t)ft}\eta \ + \ \mu_{ft} \ + \ \epsilon_{ift} \ , \end{aligned} \tag{3.1}$$

where θ_i and μ_{ft} are an individual and a field-time fixed effects, respectively, and ϵ_{ift} is the individual random productivity component assumed to be independently and identically distributed (i.i.d.) across individuals and periods.

The first step allows us to evaluate the respective explanatory power of individual characteristics, specialisation, and department-time fixed effects. The latter capture not only our observed department composition and research strategy effects but also any local effect that would be unobserved. A second step estimation allows us to identify separately the department composition and research strategy effects on the estimated department-time effect net of individual effects and selection, β_{dt}:

$$\hat{\beta}_{dt} \ = \ \text{Composition}_{dt}\gamma \ + \ \text{Department Research Strategy}_{dt}\lambda \ + \ \delta_t \ + \ \upsilon_{dt} \ , \tag{3.2}$$

where δ_t is a time fixed effect and υ_{dt} is a random component at the department level assumed to be i.i.d. across departments and periods.

The main advantage of the two-step procedure is a specification more general than one that would directly consider department variables next to individual effects in a single step, and that would ignores possible unobserved local effects. The second advantage is that the first step estimates the specialisation and individual characteristics effects independently of the specification chosen for the department composition and research strategy effects.

Changing the specification of this second step, and for instance instrumenting it or not, does not affect estimates from the first step. The two-step procedure also allows us to consider both individual and aggregated random components, which deals with the heteroscedasticity issues raised by Moulton (1990). Notice that the estimation of the second step dependent variable in the first step creates measurement error issues, which we deal with by using Feasible Generalised Least Squares (FGLS) in the second step.

The literature seems to agree on the fact that considering individual fixed effects in the first step allows the researchers to capture the role of unobserved individual effects that could otherwise bias the estimation of local or peer effects. For instance, Combes, Duranton and Gobillon (2008) show that the impact of employment density on productivity is twice lower when individual fixed effects are introduced in the specification. However to identify properly these effects separately from the location effects, one needs large data sets and enough mobility of individuals between locations. Gobillon (2004) shows that the exact identification conditions are difficult to check empirically, and it is never done in practice. Given the pretty low mobility of academics across departments and the much lower sample size by comparison with standard labour force surveys, it is difficult to be sure that individual and location effects are always properly identified. Notice also that information about age, gender and position are not often available in other studies from the literature, which makes it more important for them to control for individual fixed effects but reinforce such a concern. We present here the two sets of estimations, with and without individual fixed effects, and provide comments when conclusions differ among the two.

Introducing field-time fixed effects corresponds to an interpretation issue. This assumes that differences in publication records between fields at the World level are only a matter of fashion and size of the field but not of talents and true differences of productivity between academics and departments. Therefore, we remove them by introducing field fixed effects and focus on spatial variability independently of specialisation choices. Conversely, if one believes that the more numerous publications in a field at the world level truly corresponds to higher productivity, field fixed effects should not be introduced in the specification. We do not follow this alternative route here and introduce field fixed effects, which is also the point of view adopted in empirical economic geography that systematically considers industry fixed effects. It estimates the location effects once the composition effect due to specialisation is removed. Importantly, this does not prevent us to identify the externality role of specialisation.

Finally, one needs to comment on possible endogeneity concerns. The only way to deal with them in the first step estimation consists in using natural experiments, as proposed by Waldinger (2012) with the dismissal of scientists by the Nazi government or by Azoulay et al. (2010) with the premature death of stars. Then one has to buy that the natural experiment is not correlated with any co-variate and that estimates obtained from the natural experiment would also hold in other circumstances. Alternatively, most of the literature does not deal with such a possible endogeneity. We claim that considering both individual and department-time fixed effects (and not only a couple of department variables as often proposed) should remove most of endogeneity sources, if not all as it is the case when location choices are based on location characteristics only and not on individual temporary shocks. Since it seems to be a reasonable assumption for academics, and between two evils, we follow this strategy here.[5, 6]

Endogeneity biases can be also present in the second step estimation. For instance Combes, Duranton and Gobillon (2008) show that estimates of agglomeration economies decrease by around 20% when local variables are instrumented. Department composition and research strategy variables are endogenous when academics are mobile and have their department choices driven by the publication records of their members. Given the number of department characteristics we consider here, instrument them all is difficult and would not make much sense, particularly with respect to possible weak instrument issues. Still, as a robustness check, we show that instrumental variables estimates of models where each department characteristic is introduced alone in the specification does not change any sign of our estimates and, if anything, only increases their magnitude. We are not aware of any other paper on agglomeration and peer effects in academia that proposes an instrumentation of department characteristics.

3.2.3. Decomposing overall productivity

We emphasise that academic i productivity can be decomposed as follows:

$$y_{ift} \equiv \mathbb{1}(\text{Quantity}_{ift} > 0) \times \text{Quantity}_{ift} \times \frac{y_{ift}}{\text{Quantity}_{ift}} \tag{3.3}$$

where Quantity_{ift} is the number of publications of academic i in field f at date t. The first component is a dummy variable equal to 1 when at least one of academic i publication

5. A third strategy would consists in specifying first a model for the academic choice of department and then estimate our two equation model conditionally on that choice. However exclusion restrictions must be satisfied now, that is finding variables that explain the department choice but not the publication record. We do not see any candidate for that since even family characteristics for instance can explain the latter.

6. We also run regressions net of openness (USA and non-USA) and stars that we consider to be the most endogeneous variables of our analysis. Results are qualitatively similar and are available upon request.

refers to Jel code f. The second measures publication quantity of active academic i in field f at date t. The last component corresponds to the average quality of publications of active academic i in field f at date t. One contribution of ours consists in studying the determinants of each of these components of academics publication records separately. For instance, we can state whether a department characteristic impacts in the same direction the probability to publish, the quantity or the quality of publications, or if they are necessary substitute from each other, which is important from a policy perspective. More precisely, we assume that specifications (3.1) and (3.2) hold for each component of the individual publication record. For the first component, we estimate a logit model on the probability to publish. Then we estimate the quantity and quality determinants conditionally on publishing by using Tobit models to estimate the first step specification.[7] Moreover, in order to evaluate possible returns to scale to the number of publications on their average quality, the logarithm of quantity is introduced as an extra independent variable in the specification for quality. This also allows us to separate the direct effect of any variable on quality from its indirect effect operating through quantity, which can reinforce each other or work in opposite directions.

Finally, note that first step estimations need to weight individual observations for two reasons. First, an academic can belong to various departments. For each academic, date and field, we have as many observations as the academic's number of affiliations and each has a weight α_{idt}. Second, the academic's output is split between all the publication's Jel codes and we have, for each academic, date and department, one observation for each field

7. The inverse of Mills' ratio is calculated with a probit equation including both the individual variables and the department composition and research strategy variables. Unfortunately it is difficult to satisfy exclusion restrictions but, at least, this should allow us to control for the presence of some non-linearities in the model.

with weight $\frac{y_{ift}}{y_{it}}$. To take both effects into account, we weight by $\alpha_{idt}\frac{y_{ift}}{y_{it}}$ each observation in first step estimations.

3.3. Data

3.3.1. Measure of output

We measure the publication output of academic i in field f at date t as a weighted sum of her publications in field f listed in EconLit [8] over period τ. In most tables, τ corresponds to years $t+1$, $t+2$, $t+3$ and the output is a moving average over these three years. This choice is standard in the literature, adopted for instance recently by Ductor et al. (2011). It seems to correspond to the average reality of the profession in terms of time needed to write papers and of publication delays. [9] As a robustness check and because such a choice is both somewhat arbitrary and could result in residuals autocorrelation, we also present in Appendix 3.G estimates where τ reduces to year $t+2$. The first step results are very similar to those based on the three-year moving average. Most results of the second step are also robust to the period change: No effect changes sign and only a couple of them become non-significant, which we attribute to the noise introduced by attributing each publication to one year only.

Each publication p is first weighted by the quality of the journal, $W(p)$, where it is published. We use the Combes and Linnemer (2010) journals weighting schemes. Each journal weight is a weighted average of various recursive impact factors built from the

8. EconLit is the electronic bibliography of the American Economic Association (see http://www.aeaweb.org/econlit/index.php). It is one of the largest publication data set that lists more than $560,000$ articles published between 1969 and 2008 in more than 1200 journals.

9. Note also that the list of the department's academics at date t is establish in September of that year.

Thomson Reuters Web of Knowledge impact factors [10] and its Google Scholar citations. [11] For journals not listed by the Web of Knowledge, Combes and Linnemer (2010) use an econometric model to infer their weight. This leads to a ranking of all EconLit journals. Unfortunately, the ranking is constant overtime and all publications of a journal get the same weight independently of their publication year. We see that as a relatively minor issue that is faced in most papers in this literature. Then, a function is applied to the ranking to obtain more or less selective weighting schemes. We compare here the determinants of publications using two of them, CLm for which selectivity is moderate (it goes from a weight equal to 100 for the Quarterly Journal of Economics to a weight of 4 for the last journal, passing by 55.1 for the Journal of Labor Economics for instance) and CLh that is more selective (it goes from 100 to the Quarterly Journal of Economics to 0.0007 for the last journal, passing by 16.7 for the Journal of Labor Economics). We refer to these two schemes as the 'Quality' and 'Top quality' publication measures, respectively. They are given for the top 50 journals in Appendix 3.A.

As a common practice found in the literature, publication p is also weighted by its number of authors, $n(p)$. Since the publication output of a department is the sum of the outputs of its academics, we do not want a publication written by two members of the department to account for more (or less) than the same publication written by a single author. As mentioned above, we evaluate the presence of increasing, or decreasing, returns to scale within co-authors teams by using the average number of authors as one of the independent variable.

The third (and more minor) dimension that the output measure takes into account

10. http://www.webofknowledge.com/.
11. http://scholar.google.fr/.

is the publication's number of pages, $pa(p)$, relatively to the average page number of the publications in the journal the same year, $\bar{pa}$. This captures the idea that longer articles should contain more ideas and innovations. A natural example comes from the differences between short and regular papers in the American Economic Review. Importantly, these weights are computed within each journal-year. This assumes that the editorial policy of the journal is consistent within a year, a 20% shorter article representing 20% less output for instance. Conversely, differences in article length between journals, which can come either from different papers and font sizes or from real contribution differences, is assumed to directly and fully reflect in the journal's quality weight. In some sense, our choice is intermediate between fully ignoring the publication length and using the absolute number of pages as the literature sometimes does.

Finally, productivity is measured at the field level to enable us to study the role of fields specialisation and diversity and to control for between field differences at the world level. We use Jel codes at the first digit level (letter) and we ignore the fields "Y - Miscellaneous Categories" and "Z - Other Special Topics". We also slightly modify the codes C and D by merging code C7 (Game Theory and Bargaining Theory) and C9 (Design of Experiments) to Microeconomics (code D), which seems to us more consistent. This leaves us with 18 fields. The weight of publication p attributed to academic i is first divided by the publication's number of Jel codes, $j(p)$, and then it is multiplied by the publication's number of Jel codes corresponding to field f, $j_f(p)$.

To sum up, the publication output of academic i at date t in field f is given by:

$$y_{ift} = \frac{1}{\mathrm{Card}(\tau)} \sum_{p \in \tau} \frac{W(p)}{n(p)} \frac{pa(p)}{\bar{pa}} \frac{j_f(p)}{j(p)}$$

where Card(τ) is the number of years in period τ.

3.3.2. Academics and universities

The French Ministry of Education and Research, Cnrs and Inra [12] provided us with the list of academics in economics in France for years 1990 to 2008. Each academic is affiliated to at least one university department or to a Cnrs or Inra research centre. We merge together these affiliations at the university level to obtain what we call a "department'. This is either an economics department when it is the only affiliation where economists are found in an university (which corresponds to the majority of the cases), or the aggregation of all departments or research centres where there are economists in the university. We think that this notion of slightly aggregated economics departments better match the French reality of academic research than the detailed one. Robustness checks using detailed affiliations lead to fully consistent results, which are available upon request.

The French system allows for multiple affiliations and around 10% of academics belong to 2 or 3 departments. In this case, we give an equal weight (parameter α_{idt} in above definitions) to each department. For a few cases of academics who have positions both in France and abroad, we use their CV to evaluate the share that has to be attributed to the French department. Last, we wish to consider in the analysis only academics that can really be considered as forming a local group of academics working together. Therefore we keep only departments larger than 4 full time equivalent academics, which removes economists that are isolated in universities without real economics department. We performed a variant

12. Ministère de l'Enseignement Supérieur et de la Recherche - Direction Générale de la Recherche et de l'Innovation, Centre National de la Recherche Scientifique, and Institut National de la Recherche Agronomique, respectively. We merge by name and the initial of the firstname and then we deal with possible academics with the same name and initial manually.

keeping only departments larger than 9 full time equivalent academics and got very similar results available upon request.

The data set includes a number of individual characteristics as gender, age and position. We merge it to EconLit by surname and the initial of the first name. First names are too badly recorded in EconLit to be used integrally. This only slightly increases the number of academics with identical names, who are in any case dealt with manually. For each academic, we obtain for each year between 1990 and 2008, a data set with her individual characteristics, departments of affiliation, and publication record with weighted outputs.

3.3.3. Descriptive statistics

Table 3.1 provides the number of academics (equivalent full time) and departments per year, from 1990 to 2005. Using a three-year forward moving average for the publication output prevent us from considering the years 2006, 2007, and 2008. Both the number of academics and departments are slightly increasing over time. There are between 1,753 and 2,914 academics per year and they are affiliated to around 80 different economics departments. Over the 16 years of our panel, this leads to 38,742 academics-year observations and to 1,267 department-year observations.

Table 3.2 panel (a) presents descriptive statistics for all academics. The average academic is 45.6 year old and 25% are women. We do not present the share of each of the thirteen positions we distinguish but we create two aggregate variables that characterise them. Line 'Teaching' reports that 83% of academic-year observations have statutory teaching loads. Line 'Upper position' reports that 35% of academic-year observations correspond to an upper position, *i.e* equivalent to full professor by opposition to assistant

Table 3.1: Numbers of academics and departments per year

Year	Numbers of academics	Numbers of departments	Average department's size
1990	1753	69	25.4
1991	1853	71	26.1
1992	1933	74	26.1
1993	2038	77	26.5
1994	2175	80	27.2
1995	2292	79	29.0
1996	2365	80	29.6
1997	2423	81	29.9
1998	2530	83	30.5
1999	2680	82	32.7
2000	2724	83	32.8
2001	2744	82	33.5
2002	2752	82	33.6
2003	2764	81	34.1
2004	2803	82	34.2
2005	2914	81	36.0
Total	38742	1267	30.6

professor.

Not all academics publish over the three-year period. Line 'Publisher' of Table 3.2 panel (a) reports that one third of them do have at least one publication over the three-year period, possibly co-authored and in any field. This is one of the figures that evolved quite substantially over the period moving from 0.17 in 1990 to 0.42 in 2005. Panel (b) in Table 3.2 provides descriptive statistics on the sub-group of academics who have at least one publication over the three years. They are almost three years younger, slightly less frequently women, and they hold more frequently no teaching and upper positions.

Line 'Quantity' in Table 3.2 panel (a) reveals that the average academics writes 0.17 paper equivalent alone per year, which is one paper with one co-author every three years. This is little but partly due to the fact that many academics do not have any paper at all. Conditionally on having at least one publication over the three-year period, we

Table 3.2: Descriptive statistics

	Mean	Standard deviation	1st decile	Median	Last decile
Panel (a): All academics					
Age	45.6	9.1	32	46	58
Women	0.25	0.41	0	0	1
Upper position	0.35	0.45	0	0	1
Teaching	0.83	0.35	0	1	1
Publisher	0.33	0.44	0	0	1
Quantity	0.17	0.36	0	0	0.57
Quality	4.3	10.2	0	0	12.1
Top quality	0.80	5.31	0	0	0.22
Panel (b): Publishers					
Age	42.7	9.0	31	41	56
Women	0.22	0.38	0	0	1
Upper position	0.49	0.46	0	0	1
Teaching	0.75	0.40	0	1	1
Quantity	0.52	0.46	0.17	0.33	1.06
Quality	13.2	14.3	4.0	7.9	29.4
Top quality	2.44	8.91	0.01	0.04	4.94
Authors number	1.9	0.7	1	2	3
Non-USA openness	0.1	0.3	0	0	1.0
USA openness	0.07	0.24	0	0	0
Individual diversity	2.6	1.6	1	2	5
Panel (c): Departments					
Publishers	0.34	0.20	0.11	0.30	0.62
Quantity	5.46	8.20	0.44	2.75	12.62
Quantity per academic	0.18	0.18	0.04	0.13	0.37
Quality	11.80	8.10	5.67	9.18	20.85
Top quality	1.93	4.44	0.02	0.27	5.06
Specialisation	0.28	0.20	0.12	0.20	0.50
Size	31.6	34.6	7.5	18.0	82.0
Women	0.24	0.12	0.10	0.24	0.39
Age	45.0	3.5	40.6	45.0	49.3
Upper position	0.34	0.19	0.13	0.31	0.64
Teaching	0.79	0.34	0	0.97	1
Department diversity	-0.52	0.46	-1.18	-0.42	-0.01
Research Access	11.4	17.6	0.8	2.8	37.9
Non-USA openness	0.04	0.07	0	0.01	0.14
USA openness	0.02	0.06	0	0	0.07
Heterogeneity	2.1	0.8	1.2	1.9	3.1
Stars	0.01	0.05	0	0	0.02

Variables are defined in section 3.2. To match what is done in the econometric section, publication variables are first computed as three-year forward moving averages before descriptive statistics are computed. The number of observations for panel (a), (b) and (c) are 38,577, 12,591, and 1209, respectively. Some individuals have been deleted from Table 3.1 because of missing values. Descriptive statistics at the department level (panel (c) are calculated on the sub-sample of departments in which there is at least one publisher and hence, for which all variables are defined. That is why departments average size is slightly higher than in Table 3.1.

read in Table 3.2 panel (b) that the average number of publications is three times higher, corresponding to, for instance, one publication alone and one publication with a co-author every three years. As regards quality we also confirm the large disparities existing among academics, a well-documented fact since Lotka (1926). The median publication is worth an equivalent of a publication in the 150th journal but the median is lower, around the 350th journal. By contrast, the top decile average quality publication corresponds to a publication in the top 50th journal. The average quality of publications of academics in France appear to be relatively better in terms of the top quality index since the mean is now around the top 50th journal, the median around the top 100th, and the top decile around the top 30th journal. 10% of the publishers have at least one co-author abroad but not in the USA, and 7% have at least one co-author in the USA. The average number of authors per paper is 1.9 but more precisely there are 44.7%, 38.0% and 14.8% of the publications with one, two, and three authors, respectively. Only 2.5% of the publications have strictly more than three authors.

Regressions are performed at the field level. Since there are 18 possible fields, the 38,577 academics-year observations translates in 694,386 field-academic-year observations, which is then increased by the fact that some observations are duplicated for those academics with multiple affiliations as explained in section 3.2.2. As a result, the number of observations we have in the first step estimation for the probability to publish is 771,426. However, both because some academics do not publish at all, and, most importantly, none publishes in all fields, many of these observations corresponds to zero publications (in the field). There are 'only' 38,984 non-zero observations, which are the observations for the first step quantity and quality estimations. Line 'Individual Diversity' in Table 3.2 panel (b) reveals that the average number of fields per academic over a three-year period is 2.6 and the very diversified academic at the top decile has five of them. At the national level, "Microeconomics' is largely the most represented field in France with 16.8% of the publications number. This is larger than its share for EconLit as a whole, which is 10.2%. Then, there are ten fields representing more than 4%. [13]

Panel (c) in Table 3.2 reports descriptive statistics at the department level. The average department has 31.6 academics who are 45 year old on average, 24% are women, 34% have upper positions, and 34% are publishers. The figures are comparable to the averages over all academics but they all present quite a lot of variations between departments, which is also observed for publication records. The average department has a stock of 5.5 publications per year, 0.18 per academic, and the average quality indexes are in the same ranges as for individual academics. The median specialisation means that a Jel code represent on average 20% of the department publications, which is high since there are 18 possible different Jel codes. In the very specialised department at the top decile of specialisation, each Jel code represents half of the publications, that is there are only two Jel codes represented within the department. This is confirmed by the diversity index, which takes almost always negative values, meaning that departments are less diversified than what would result from random Jel codes choices.

Finally, Table 3.3 presents the simple correlations between the variables at the department level. First, quantity and quality are largely positively associated even for the top

13. Industrial Organization (9.5% vs 8.8% for EconLit as a whole), Development/Growth (8.8 vs 10.0%), Finance (8.8 vs 10.9%), Macro/Monetary Economics (8.2 vs 7.2%), Labour/Demography (8.2 vs 8.3%), International Economics (7.6 vs 7.8%), Agricultural/Environmental Economics (5.6 vs 7.0%), Economics History, Thoughts and Methodology (5.4 vs 2.2%), Public Economics (4.2 vs 4.3%), Urban and Regional Economics (4.2 vs 5.0%).

Table 3.3: Simple correlations at the department level

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Quantity (1)	0.92	0.75	0.08	-0.07	-0.08	0.48	-0.36	0.59	0.32	0.55	0.48	-0.78	0.42
Quality (2)	1	0.92	0.11	-0.08	-0.09	0.52	-0.43	0.52	0.37	0.63	0.59	-0.69	0.54
Top quality (3)		1	0.21	-0.07	-0.06	0.48	-0.38	0.43	0.36	0.57	0.55	-0.51	0.49
Size (4)			1	-0.02	0.19	0.06	0.15	0.16	-0.09	0.03	-0.04	0.16	-0.09
Women (5)				1	-0.08	-0.28	0.16	0	0.23	-0.07	-0.07	-0.02	-0.15
Age (6)					1	0.26	0.03	0.03	0.23	-0.15	-0.11	0.19	-0.14
Upper position (7)						1	-0.58	0.15	0.40	0.38	0.43	-0.37	0.40
Teaching (8)							1	-0.05	-0.36	-0.37	-0.46	0.41	-0.36
Diversity (9)								1	0.19	0.25	0.18	-0.50	0.15
Research Access (10)									1	0.30	0.34	-0.31	0.30
Non-USA Openness (11)										1	0.61	-0.45	0.51
USA Openness (12)											1	-0.39	0.72
Heterogeneity (13)												1	-0.29
Stars (14)													1

Variables are defined in section 3.2. For specialisation that is defined at the Jel code level, it is first averaged by department (weighted by the share of the Jel code in the department) before statistics are computed.

quality index. Those departments that publish more also produce higher quality publications and no trade-off seems to take place between the two. This is reminiscent to what Combes and Linnemer (2003*b*) find at the European level. Academics are also on average more productive in departments where the share of upper positions is higher and the share of teaching positions lower, where field diversity and research access are high. Correlations are also positive and large with the share of department academics connected to co-authors abroad and in the USA (openness variables), for the presence of stars and heterogeneity that are positively and negatively correlated with quantity and quality, respectively. The correlation of size with quantity is not very large but it increases for quality, and even more for top quality. We have now to investigate whether these correlations are driven by the fact that upper position researchers, or researchers with high abilities more generally, are over-represented in some departments through selection effects and/or by the fact that some academics or some department characteristics generate more externalities. This is the purpose of the econometric analysis developed in the next sections.

3.4. Productive individuals: Abilities and research strategy versus location

This section studies the determinants of individual productivity and assesses the relative weight of individual and department effects, respectively. We regress individual productivity in a specific field on individual characteristics that relate to both individual abilities and individual research strategy (including field-time fixed effects), department specialisation and department-time fixed effects. Columns (1) and (2) ('Publishing') in Table 3.4 regard a Logit model where the dependent variable is 1 if academic i produces in field f at date t and 0 otherwise. Columns (3) and (4) ('Quantity') regard the number of publications, Columns (5) and (6) ('Quality') and Columns (7) and (8) ('Top quality') the publication average quality using the regular and a top journal quality indexes, respectively. For each output measure, the first column does not include the department variables (specialisation and the department-time fixed effect), which are included in the next column specification. Table 3.17 in Appendix 3.B reproduces Table 3.4 including individual fixed effects.

Before moving to the effect of each variable, let us start with some variance analysis. A first conclusion is that the model better explains the average quality of publications than the number of publications, even more when a top quality index is considered and when individual fixed effects are introduced. The publication quality relates more to individual and department characteristics than the number of publications for which the random component is larger. This would probably make sense to any academics since publishing

Table 3.4: Determinants of individual publications

	Publishing		Quantity		Quality		Top quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Women	-0.368[a]	-0.373[a]	-0.090[a]	-0.376[a]	0.043[a]	-0.297[a]	0.102[b]	-1.024[a]
	(0.02)	(0.01)	(0.01)	(0.04)	(0.01)	(0.04)	(0.04)	(0.09)
Age	-0.104[a]	-0.090[a]	-0.030[a]	-0.107[a]	0.002	-0.084[a]	-0.010	-0.292[a]
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Age square	0.000[a]	0.000[a]	0.000[a]	0.001[a]	-0.000	0.000[a]	0.000	0.001[a]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number			-0.917[a]	-0.868[a]	0.292[a]	0.279[a]	1.050[a]	0.989[a]
			(0.01)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
Non-USA Openness			0.362[a]	0.319[a]	0.253[a]	0.242[a]	0.889[a]	0.854[a]
			(0.01)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
USA Openness			0.338[a]	0.322[a]	0.408[a]	0.377[a]	1.243[a]	1.129[a]
			(0.02)	(0.01)	(0.03)	(0.01)	(0.07)	(0.04)
Star			0.492[a]	0.413[a]	0.877[a]	0.772[a]	2.504[a]	2.156[a]
			(0.03)	(0.02)	(0.04)	(0.02)	(0.12)	(0.06)
Diversity			-0.131[a]	-0.060[a]	-0.002	-0.007	0.087[a]	0.074[a]
			(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
Specialisation				0.368[a]		0.007		0.024[c]
				(0.00)		(0.01)		(0.01)
Quantity					0.087[a]	0.075[a]	0.491[a]	0.442[a]
					(0.01)	(0.01)	(0.02)	(0.01)
Selection			-0.180[a]	1.861[a]	-0.565[a]	1.662[a]	-1.743[a]	5.600[a]
			(0.04)	(0.23)	(0.05)	(0.23)	(0.14)	(0.61)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Jel Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department Time FE	No	Yes	No	Yes	No	Yes	No	Yes
R^2			0.284	0.425	0.330	0.398	0.433	0.501
Observations	771426	760122	38984	38984	38984	38984	38984	38984

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. 'Publishing': Logit model for the probability to have at least one publication. 'Quantity', 'Quality', and 'Top quality': Heckman two-step Generalized Tobit models. The first step consists in a Probit model for the probability to publish. The variable 'Selection' in the second step is the inverse of Mills' ratio from the Probit equation. Variables are defined in section 3.2.

in good journals needs more specific skills, that would be captured by the model, than just publishing.

Importantly, there is a large increase of the R^2 when department effects are introduced, by almost 50% for the quantity published, 21% for the publication quality and 16% for top quality (as regards estimations without individual fixed effects). This is confirmed by the more detailed variance analysis provided in Table 3.5. First, columns 'Stand. error' report the standard error of the effect of a variable or of a group of variables for the quantity estimation presented in Table 3.4 column (4) and for the same regression including individual fixed effects presented in Table 3.17 column (4) in Appendix 3.B. The higher it is by comparison with the standard error of the dependent variable to be explained reported in the first line, the larger the explanatory power of this variable or group of variables. However, and most importantly, a variable or a group of variable, has a large explanatory power when its effect is largely correlated with the dependent variable. This is reported in columns 'Correlation'.[14] We observe that the standard error of individual effects is slightly larger than the standard error of department effects. By contrast, the correlation of the latter with the dependent variable is slightly higher than for the former. Both results point to the fact that the two groups of variables have a relatively similar explanatory power of the individual number of publications, at least when individual fixed effects are not considered.

This shows for such a type of analysis that considering or not an individual fixed effect in the specifications is crucial. First, the explanatory power of the model largely increases when individual fixed effects are introduced. It almost doubles in the regressions without department effects, and even when those are present, it is 50 to 80% higher than in the model without individual fixed effects (see Table 3.17 in Appendix 3.B). The explanatory power of the model is now comparable, even if a bit lower, to what is obtained in standard individual wage equations, with R^2 between 0.60 for quantity and 0.73 for top quality. Second, from the right-hand-side of Table 3.5, we observe that the standard error of individual

14. The variance analysis is computed on variables centered with respect to their annual means and therefore are performed in the within time dimension, to focus on spatial variations. See Abowd, Kramarz and Margolis (1999) for details on this type of variance analysis.

effects is now twice larger than the one of department effects while the correlation with the dependent variable is also larger for the former. This means that some unobserved individual effects do significantly influence the number of publications, which increases the explanatory power of individual effects relatively to department effects.

Tables 3.18 and 3.19 in Appendix 3.C reproduce the variance analysis for the publication quality and top quality, respectively. We first observe that individual effects explain more quality than quantity. Even when individual fixed effects are not included, the explanatory power of individual effects is larger than that of department effects. At the extreme, department effects have a standard error and a correlation with the dependent variable around three times lower than those of individual effects.

To sum-up, and keeping in mind the discussion in section 3.2.2 about the fact that individual fixed effects cannot always be properly identified separately from department effects if there is not enough mobility between departments which could be the case here, the lower bound for the explanatory power of department effects would be around half of the explanatory power of individual effects. However, at the upper bound, without individual fixed effects but still a bunch of individual observable characteristics, department effects could explained as much as individual effects, even more for the number of publications than for their average quality. This means that agglomeration and peer effects do matter for individual productivity in academics. Interestingly, the lower bound is of a similar magnitude to what Combes, Duranton and Gobillon (2008) find for market activities in the same country, France. Individual effects would not play a larger role, and possibly a lower one, in scientific activities relatively to the role of location by comparison with other activities. This contrasts with the findings of the literature. Waldinger (2012) does not find any peer effects among physicists, chemists, and mathematicians in Nazi Germany, which is also the conclusion of Dubois et al. (2011) for Mathematicians nowadays. Kim et al. (2009) find that the effect of being in a top 25 economics and finance department progressively disappear between the 1970s and the 1990s in the USA. First, notice that these authors comment the fact that department effects are, or not, significant but do not discuss their global explanatory power, which we do here. This is not fully the same point of view and we find our approach more relevant to assess the share of individual productivity explained by location. Then, another possible explanation of the difference between these results is that individual fixed effects are not always properly identified. Unfortunately, as discussed in section 3.2.2 the fact that mobility is high enough to identify both individual and location fixed effects is difficult to formally test. A last explanation could be that research habits would differ between an European country as France and the USA, both in terms of research technology (as the intensity of the Internet use for collaborations for instance) and in terms of institutional design. For instance the possibility of individuals to capture their publication performance is considered to be lower in most European countries where wages and positions much less relate to publication records. [15] All of this could affect the relative role of individual and department characteristics.

Another crucial result emphasised by Combes, Duranton and Gobillon (2008) regards the sorting of workers across space. More able workers locate in more favorable locations, those where location effects reflecting localised externalities are the largest. From the econometric point of view, it is important to assess whether department effects would be biased if individual effects were ignored. From the policy point of view, it is interesting to

15. Combes, Linnemer and Visser (2008) document it for France.

Table 3.5: Variance analysis of the individual publication quantity

	Without individual fixed effects			With individual fixed effects		
	Stand. error	Correlation	Sorting	Stand. error	Correlation	Sorting
Explained: quantity	0.456	1.000		0.456	1.000	
Individual effects	0.439	0.335	0.100	0.497	0.484	0.029
Individual fixed effect	-	-	-	0.265	0.460	-0.008
Gender	0.086	0.094	0.014	-	-	-
Age	0.295	0.023	-0.053	0.110	0.027	-0.029
Position fixed effect	0.315	0.091	0.126	0.186	0.096	0.056
Authors number	0.187	0.353	-0.025	0.184	0.353	-0.018
Non-USA openness	0.063	0.110	0.106	0.034	0.110	0.088
USA openness	0.047	0.118	0.116	0.029	0.118	0.098
Star	0.038	0.168	0.157	0.027	0.168	0.135
Individual diversity	0.019	0.046	-0.162	0.030	0.046	-0.156
Jel fixed effect	0.201	0.087	0.042	0.272	0.073	0.029
Department effects	0.269	0.372	0.743	0.250	0.333	0.735
Department fixed effect	0.246	0.155	1.000	0.226	0.112	1.000
Specialisation	0.186	0.333	-0.248	0.175	0.333	-0.240
Selection	0.436	-0.134	-0.396	0.441	-0.124	-0.334
Residuals	0.349	0.766	0	0.291	0.640	0

The table presents the variance analysis of the estimation reported in Table 3.4 column (4) and of the same regression including also individual fixed effects that is reported in Table 3.17 column (4) in Appendix 3.B. All variables are first centred with respect to their annual mean. Columns 'Stand. error' report the standard error of the effect of a variable or of a group of variables. For the first line, it reports the standard error of the dependent variable. Columns 'Correlation' reports the correlation between the effect of a variable or of a group of variables with the dependent variable. Columns 'Sorting' reports the correlation between the effect of a variable or of a group of variables with the department fixed effect.

know whether more productive academics are attracted by the departments that generate more externalities, or if they locate randomly across departments. Columns 'Sorting' in Table 3.5 for publication quantity, and in Tables 3.18 and 3.19 in Appendix 3.C for the average quality and top quality of publications, report the correlation between the variable or group of variables and the department fixed effects. It is typically found to be positive for individual effects, which means that workers with individual characteristics that make them publishing more and with a higher quality are located in the departments that provide larger external effects. The correlation of all individual effects together, at 0.1, is significantly lower than what Combes, Duranton and Gobillon (2008) find for market activities in France. However it is larger for the publication quality, and even larger for top quality publications. Due to the presence of such a correlation, Combes, Duranton and Gobillon (2008) show that not considering individual effects can largely bias the estimation of department effects, which we will also illustrate below. Interestingly, we find that the spatial sorting of academics is larger on observed characteristics than on unobserved ones since the correlations are smaller when individual fixed effects are considered (right-hand-side part of the tables). However, the gap between the two reduces for quality, and especially top quality, meaning that unobserved characteristics matter relatively more when the quality of publications increases. This is again consistent with intuition since it is *a priori* less easy to identify the good mix, as well in terms of individual abilities as in terms of research strategy, to publish in the best journals. Unfortunately, none of the papers assessing the magnitude of peer effects in science compute such correlations between individual and department effects, which would allow us to compare them with those in other fields or for other periods.

We finally turn to the role of each variable. From Table 3.4, women and older academics appear to publish less, at least when department effects are considered. This is consistent with previous findings of the literature, even more when one recalls that we control here for the type of position hold. Once a given position is achieved, as for instance having a full professor rank, the number and quality of publications decreases with age. Part of the effect might also result from a cohort effect (previous generations had less strong incentives to publish than younger academics), which would be consistent with the fact that the impact of age is not significant anymore when individual fixed effects are considered.

As detailed in Table 3.20 in Appendix 3.D, expected results are obtained for the impact of the various positions. The higher the rank (professor, research professor, and even more

Insee or Ponts Engineers by opposition to assistant professors or research fellows) and the larger the time allocated to research (research versus teaching positions), the larger the published quantity and its quality, which is also the case for the academics purely in economics. Therefore, even if part of promotions in France does not relate to publications as emphasised by Combes, Linnemer and Visser (2008), those who get better positions do publish more on average. Note that our purpose here is not to give a causal interpretation to such variables but to control at best for individual abilities when estimating the role of departments.

Interestingly, we do not control only for some of the standard 'ability' variables considered in wage or productivity equations, as gender, age or position (which plays the role of occupation), but we also consider variables related to what we call the individual 'research strategy'. The variable that has the effect with the largest correlation with the dependent variable is the academic's average number of co-authors per publication. Its impact on published quantity is largely negative. Having more co-authors decreases the number of published papers, which means that attributing only part of the publication to each co-author corresponds to a stronger effect than one should also produce more papers with more co-authors. In other words, there are for the number of authors decreasing returns to scale for the quantity published, academics would publish more papers if they would work alone. However, the average number of co-authors has a large positive effect on the average publication quality, which is even larger for top quality. Therefore, a larger number of co-authors decreases the number of publications equivalent written alone but increases their quality. There is a trade-off between the two, and only an analysis such as this one allows to identify the two separately. For instance, an academic that has on average two co-authors instead of one only (per publication) has 30% less publications but their average quality is 12% higher and their average top quality is 49.3% higher.[16] This is from the estimations without individual fixed effect but those considering it lead to very similar magnitude. Having two co-authors instead of one decreases the quality weighted volume of publications (a measure frequently found in the literature that consists here in multiplying the quantity by the average quality) by around 21.6% but increases the top quality weighted volume of publications by around 4.5%. Therefore, one should not expect to publish more, even in decent journals, thanks to co-authoring but possibly to reach top journals. Notice that the number of co-authors also has an indirect negative effect on average quality through quantity since quantity has a positive impact on average quality (which is commented below). However, this indirect effect is not large enough to compensate the positive direct effect on average quality. The total effect of co-authorship on quality is at 0.227 and on top quality at 0.605.[17] Sauer (1988) who is one of the earliest contributions on the role of co-authorship finds almost no effect, and two other studies, also regarding economists, Hollis (2001) and Medoff (2003) conclude to a negative effect of co-authorship on publication quality. Dubois et al. (2011) identify an overall negative effect of co-authorship for mathematicians on their citation-weighted publication index but it turns positive when the co-author specialisation is taken into account. The difference between these studies' results and ours can be due to the fact that they do not precisely distinguish the quantity and quality effects, which go in opposite directions as we show. Consistent with our result is also Ductor (2011) who finds a negative effect of co-authorship for economists between 1971 and 1999 that turns positive when unobserved heterogeneity

16. $1.5^{-0.868} - 1$, $1.5^{0.279} - 1$ and $1.5^{0.989} - 1$, respectively.
17. $-0.868 \times 0.075 + 0.292$ and $-0.868 \times 0.442 + 0.989$, respectively.

and endogenous co-authorship formation are taken into account. [18]

We also find that having a higher diversity of research fields does not really help academics to publish more but it increases their average publication quality, and again the effect is larger for top quality. This suggests the presence of some complementarities between research in different fields at the individual level. Dubois et al. (2011) also find a positive effect of field diversity for mathematicians.

Finally, estimations for quality control for quantity, the individual number of publications. This allows us to test for the presence of increasing returns to scale for quality at the individual level now (by opposition to the co-author team level assessed through the number of co-authors). This is not done in the literature usually. Table 3.4 reveals that there are indeed increasing returns to the number of publications for average quality, and even more for top quality publications. The more academics publish, the higher the average quality of their publications. An academic with twice more publications has an average publication quality higher by 5.3% and a publication top quality higher by 35.8%. [19]

Notice that all these results holds within Jel codes since we control for Jel codes fixed effects. Jel codes appear to have a pretty large explanatory power especially as regards publication quality. This reflects that all fields are not equal in terms of publication opportunities. Now, to the best of our knowledge, nobody has been able to assess yet whether this is due to a pure 'fashion' effect (some topics are more in fashion, which makes them easier to publish) or to some selection effects (more able academics self-select in some fields or some fields attract more able academics). However, it is not the purpose of the present article to tackle this difficult question. Still, in terms of interpretation, departments effects

18. See section 3.2.2 for a discussion about endogeneity concerns in the first step estimation.
19. $2^{0.075} - 1$ and $2^{0.442} - 1$, respectively.

are estimated net of the direct role of the academic composition in terms of research fields, which is what matters here.

3.5. Productive departments: Sorting versus externalities

The previous section studies how much location matters for the individual productivity of academics. We now move to the dual question that assesses the extent to which the department composition in terms of individual characteristics explains the department performance relatively to the presence of local externalities. Therefore we repeat the previous variance analysis but after aggregating variables by department. Table 3.6 presents the results for quality while Tables 3.21 and 3.22 in Appendix 3.E do it for quantity and top quality, respectively.

Both Kim et al. (2009) for Economics and Business and Dubois et al. (2011) for Mathematics argue that what makes a good department nowadays is primarily gathering academics with good individual characteristics. This is not what we find here. Even according to the specification that includes individual fixed effects, in the right-hand side of Table 3.6, the standard error of individual effects is only slightly higher than the standard error of department effects. Still, the correlation with the department average quality of publication is twice higher for the former. When individual fixed effects are not controlled for, both the standard error of department effects and their correlation with the dependent variable are now higher than those for individual effects. Therefore, at best, the characteristics of academics explain a bit more the publication quality disparities between departments than the external effects at stake in these departments. From Tables 3.21 and 3.22 in Appendix 3.B, the same conclusion is achieved for quantity and top quality, to slightly larger

Table 3.6: Variance analysis of the average publication quality at the department level

	Without individual fixed effects			With individual fixed effects		
	Stand. error	Correlation	Sorting	Stand. error	Correlation	Sorting
Explained: quality	0.426	1.000		0.426	1.000	
Individual effects	0.494	0.532	0.225	0.536	0.721	0.060
Individual fixed effect	-	-	-	0.383	0.694	0.022
Gender	0.059	0.054	0.005	-	-	-
Age	0.266	0.076	-0.119	0.089	0.072	-0.018
Position fixed effect	0.300	0.324	0.264	0.124	0.104	-0.047
Authors number	0.068	0.358	0.096	0.061	0.358	0.134
Non-USA openness	0.040	0.414	0.184	0.010	0.414	0.099
USA openness	0.038	0.484	0.315	0.018	0.484	0.240
Star	0.058	0.514	0.348	0.044	0.514	0.222
Individual diversity	0.002	-0.239	-0.383	0.001	0.239	0.335
Jel fixed effect	0.216	0.246	0.083	0.205	0.209	0.036
Department effects	0.524	0.712	1.000	0.381	0.495	1.000
Department fixed effect	0.526	0.711	1.000	0.382	0.494	1.000
Specialisation	0.004	-0.110	-0.426	0.005	-0.110	-0.335
Quantity	0.028	0.118	0.093	0.021	0.118	-0.038
Selection	0.528	-0.404	-0.635	0.386	-0.391	-0.522
Residuals	0.001	-0.050	0	0.001	-0.037	0

The table presents the variance analysis of the estimation reported in Table 3.4 column (6) and of the same regression including also individual fixed effects that is reported in Table 3.17 column (6) in Appendix 3.B, once they are averaged by department. For the meaning of the figures reported, see the footnote in Table 3.5.

and smaller extents, respectively. Again, this means that agglomeration and peer effects are quite strong and therefore important to explain the ranking of academic institutions.

Among individual characteristics that do explain department disparities, research fields present in the department play a large role, with a standard error and a correlation with the dependent variable of Jel code fixed effects at 0.216 and 0.246, respectively. Having a larger share of high rank positions is the only group that has a larger explanatory power. Importantly for policy implications, the share of each position type is for a large extent not a department decision in France but results from individual location choices for research positions, and from the Ministry of Higher Education decisions for positions with teaching. It also appears that the higher publication quality in some departments is due to the fact that they gather more people with a higher average number of co-authors. Having academics with co-authors abroad or being a star is also pretty correlated with the department average publication quality but the explanatory power of these variables is reduced by their pretty low variability across departments. Individual diversity does not vary so much between departments and its effect is not strongly correlated with the department average quality even if its elasticity is largely positive. Differences in gender and age composition does not really matter for disparities between departments. As presented in Appendix 3.E, such conclusions are broadly confirmed for quantity and top quality with some variables having sometimes a slightly larger or lower explanatory power.

We can also study the impact of the individual spatial sorting on department disparities. For market activities in France, Combes, Duranton and Gobillon (2008) find that the correlation between individual and department fixed effects is large, at 0.29. Disparities between individual characteristics and in terms of local externalities therefore cumulate and generate pretty large productivity disparities between locations. Individual and department effects disparities are less systematically related to each other here since the correlation between both is at 0.06 only, as appears in the right-hand side of Table 3.6. Interestingly, sorting on individual observed characteristics, which is reported in the left-hand side of Table 3.6, is larger, at 0.225. This means that unobserved individual characteristics, that significantly increase the explanatory power of the model (the correlation of individual effects with the dependent variable is 0.72 against 0.53 without individual fixed effects), are distributed more independently from department effects than observed characteristics.

Kim et al. (2009) argue that the declining role of economics and finance department externalities is a recent trend that arose progressively in the eighties and then in the nineties, by comparison with the seventies. To assess if such a trend is also present for French economics departments, and because our panel spans over a pretty long period of time, we repeat all our analysis for two sub-periods separately, 1990–1997 and 1998-2005. These periods are interesting because it is only at the end of the nineties that, first, the use of Internet for literature search and papers diffusion starts being systematised in France and, second, that publications in peer-reviewed non-French journals became the norm when evaluating academics. Both could have contributed to a change in the role of departments for publications. However, all the conclusions we draw in this paper are extremely stable between the two-periods and correspond to what is found over 1990-2005. In particular we do not observe any decline in the strength of department externalities over time. Results are available upon request.

3.6. The channels of department externalities

The last step of our analysis consists in evaluating the channels through which department externalities operate. This consists in studying both the impact of the specialisation variable in the first step of the estimation and the determinants of the department fixed effect in the second step estimation.

Specialisation is the only variable that is both department and Jel code specific. It assesses whether having a large share of department publications in a field helps academics to publish in this field. This is called a localisation effect in economic geography,[20] which can reflect external (to the firm or individual but local) economies of scale taking place

20. For all references to empirical economic geography in this section, please refer to Combes, Mayer and Thisse (2008)

within industries, fields here. Indeed, Table 3.4 shows that department specialisation has a positive effect on the quantity published in the field by an academic. The effect, at 0.368, is pretty large since the specialisation elasticity is usually found in the range 0.010-0.050 for productivity in market activities. This is somewhat counterbalanced here because specialisation variability is lower across departments than specialisation across cities. Still, when one increases the share of publications in a field in the department by 50% (which corresponds to half a standard error at the median), the number of papers published in the field increases by 16.1%.[21]

By contrast, there is no direct significant impact of specialisation on the quality of publications, be it measured with the medium or top journal quality index. Still, specialisation has an indirect positive impact on quality due to the positive impact of quantity on quality. The indirect impact of specialisation is at 0.027 for quality and 0.163 for top quality.[22] Last, and by contrast to many other determinants of department fixed effects, the impact of specialisation is almost not affected when individual fixed effects are introduced in the specification (Table 3.17 in Appendix 3.B).

Table 3.7 reports the impact of department variables on the estimated department fixed effect for each publication variable. Table 3.24 in Appendix 3.F does it when individual fixed effects are controlled for in the first step estimation. Since the dependent variable is estimated in a first step, one needs to correct for measurement errors on it. We do so using Feasible Generalised Least Square (columns 'FGLS') and we systematically compare the estimates with Ordinary least Squares (columns 'OLS'). Table 3.7 shows that results are almost not sensitive to the use of OLS or FGLS. This confirms that department fixed

21. $1.5^{0.368} - 1$.
22. 0.368×0.075 and 0.368×0.442, respectively.

Table 3.7: Determinants of department fixed effects

	Publishing		Quantity		Quality		Top quality	
	OLS (1)	FGLS (2)	OLS (3)	FGLS (4)	OLS (5)	FGLS (6)	OLS (7)	FGLS (8)
Size	0.100[a]	0.101[a]	0.198[a]	0.174[a]	0.067[a]	0.078[a]	0.172[a]	0.192[a]
	(0.014)	(0.014)	(0.012)	(0.012)	(0.013)	(0.013)	(0.033)	(0.034)
Women	0.230[b]	0.229[b]	0.490[a]	0.355[a]	0.334[a]	0.448[a]	1.173[a]	1.558[a]
	(0.102)	(0.104)	(0.090)	(0.109)	(0.095)	(0.114)	(0.247)	(0.297)
Age	0.034[a]	0.033[a]	0.029[a]	0.021[a]	0.025[a]	0.022[a]	0.081[a]	0.072[a]
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.010)	(0.013)
Diversity	0.662[a]	0.662[a]	0.448[a]	0.367[a]	0.323[a]	0.329[a]	1.037[a]	1.019[a]
	(0.026)	(0.026)	(0.023)	(0.046)	(0.024)	(0.047)	(0.063)	(0.123)
Research Access	−0.017[b]	−0.017[b]	−0.011	−0.005	0.015[b]	0.013[c]	0.042[b]	0.031
	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.020)	(0.020)
Non-USA Openness	1.287[a]	1.290[a]	1.451[a]	1.262[a]	1.435[a]	1.347[a]	4.941[a]	4.680[a]
	(0.202)	(0.202)	(0.179)	(0.220)	(0.188)	(0.229)	(0.491)	(0.597)
USA Openness	0.858[a]	0.853[a]	0.210	0.088	1.023[a]	1.084[a]	3.522[a]	3.563[a]
	(0.276)	(0.276)	(0.245)	(0.228)	(0.257)	(0.241)	(0.669)	(0.628)
Heterogeneity	−1.147[a]	−1.145[a]	−1.082[a]	−0.900[a]	−0.752[a]	−0.757[a]	−2.539[a]	−2.493[a]
	(0.039)	(0.039)	(0.035)	(0.105)	(0.037)	(0.107)	(0.095)	(0.280)
Stars	0.828[b]	0.832[b]	1.549[a]	1.387[a]	2.527[a]	2.538[a]	7.691[a]	7.759[a]
	(0.375)	(0.375)	(0.332)	(0.328)	(0.349)	(0.345)	(0.909)	(0.899)
Positions	yes	yes	yes	yes	yes	yes	yes	yes
Time Fixed Effect	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.82	0.82	0.86	0.93	0.75	0.85	0.82	0.90
Observations	1209	1209	1209	1209	1209	1209	1209	1209

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively.

effects are pretty precisely estimated in the first step, or at least that no large bias could result from measurement error on them. When individual fixed effects are considered, as presented in Table 3.24, the same conclusion is reached.

The impact of the size of the local economy on local productivity is one of the most studied question in economic geography. One could have controlled for such a variable here as the total size of the city where the university is located for instance. However, we think that local externalities can be even more localised as regards academic activities that need face to face contacts. Therefore, we use the size of the department defined as its number of academics, which is also a variable interesting *per se* since at least partially in the hands of the department head, of the university or of the central gouvernement (in many European countries for instance). We then test the relevance of our choice of spatial scale in two ways. First, we also include a research access variable that corresponds to the proximity to other departments, which allows us to disentangle very local externalities from more extended ones. Second, we provide estimates in section 3.7 at the city level.

Department size has a positive and significant impact on all measures of individual productivity: the probability to publish, the number of publications and their average quality. The largest effect is obtained for the average quantity of publications, at 0.198.[23] It is much larger than standard estimates in economic geography for city size that are at most at 0.080. We obtain, as in the economic geography literature, that the impact much decreases when individual fixed effects are controlled for in the first step estimation, even if we saw above that spatial sorting was less marked for academics than for market activities. Still, the effect is now 0.161 for quantity (OLS estimate in Table 3.24), still much larger than the standard 0.020 found for market activities when individual fixed effects are controlled for, and even if the effect is not significant anymore for the other dimensions of publication. When individual fixed effects are controlled for, doubling the size of a department (which corresponds to around half a standard error at the median) increases the average quantity of articles by 11.8%,[24] which has in turn an extra positive indirect effect on quality and top quality. The impacts are positive and significant when individual fixed effects are not controlled for. From the variance analysis reported in Table 3.8 for quality and in Tables 3.25, 3.26 and 3.27 in Appendix 3.F for publishing, quantity and

23. We use the OLS estimates here that are more directly comparable with the literature but we already mentioned the fact that in most cases differences with FGLS are small, unless for quantity when individual fixed effects are controlled for, which prevent from using FGLS.

24. $2^{0.161} - 1$.

top quality, respectively, we observe that department size has some explanatory power of the department fixed effect, especially for quantity. However many other variables, commented below, have a larger explanatory power. This contrasts with the usual findings of the economic geography literature where size is found to be the main explanation of productivity differences across locations. Therefore larger departments do make academics more productive but other factors can play an even larger role.

Research access has little significant impact on department externalities.[25] This contrasts with the economic geography findings where market access variables present almost always large and positively significant effects on productivity. Therefore, agglomeration effects appear to be very localised for academic activities, more than for market activities in general. This finding is consistent with an important role for face to face contacts for some specific activities as argued by Gaspar and Glaeser (1998). For instance Arzaghi and Henderson (2008) find for the advertising agency industry that agglomeration effects take place at the block level in Manhattan. The very localised nature of interactions in science is also consistent with Agrawal and Goldfarb (2008) who find that bitnet (an early version of internet) did not impact more the academic collaborations between very distant institutions. Once outside the department, the impact of other academics does not depend on where they are located.

It is important to notice that we control, at least partly, for the role of co-authors through the openness variables. Distance to co-authors is shown by the literature on academic networks (see for instance Laband and Tollison (2000), Rosenblat and Mobius (2004) or Goyal, van der Leij and Moraga-González (2006)) to have significantly increased

25. Alternative specifications of the research access variable, with squared distance or square root of distance in the denominator leads to qualitatively similar results.

over time. Would the links to co-authors not be controlled for, research access could have had a stronger effect, at least for the first years in our sample for which the internet use was less extended. Not only having co-authors abroad and in the USA increases both the individual quantity and quality of publications for an academic as we show in section 3.4 but we also find that a larger share in the department of academics with co-authors abroad creates a positive publication externality for all academics, as appears from the lines 'Non-USA openness' and 'USA openness' in Table 3.7. We probably capture that in a world where distance does not matter once outside the department, being connected with other academics elsewhere, and in particular in the USA where a large share of academic activity takes place, is important. This is fully consistent with the large role of networks in academia that is underlined by the literature we have just quoted. We find that all dimensions of the publication activity (probability to publish, number of publications and their quality) are affected but the effect is larger and larger when one moves across these dimensions. Still, these two variables do not have a very large explanatory power since the variance of their effect is lower than that of department size while the correlation with the department fixed effect is similar, as reported in Table 3.26.

Similarly, both women and older academics do exert positive externalities on other academics' publications, again whatever the publication dimension, but these variables have an even lower explanatory power of the department fixed effect than department size and openness. Therefore, these are not the effects that drive externality disparities between departments. Still, it is interesting to notice that two categories of academics who publish less individually as we show in section 3.4 exert a positive externality on the publications of their colleagues. Also interesting is the fact that the reverse is observed

Table 3.8: Variance analysis of the department fixed effects for the average publication quality

	Without individual F.E.		With individual F.E.	
	Stand. error	Correlation	Stand. error	Correlation
Explained: department fixed effect	0.525	1.000	0.381	1.000
Composition effects	0.207	-0.129	0.144	-0.084
Size	0.058	0.055	0.017	-0.022
Gender	0.036	-0.088	0.040	-0.023
Age	0.086	-0.016	0.068	-0.068
Positions	0.153	-0.164	0.120	-0.051
Research strategy effects	0.510	0.760	0.313	0.608
Diversity	0.139	0.537	0.105	0.465
Research Access	0.023	0.303	0.024	-0.135
Non-USA Openness	0.093	0.473	0.033	0.292
USA Openness	0.059	0.453	0.055	0.318
Heterogeneity	0.273	0.678	0.183	0.563
Stars	0.122	0.409	0.057	0.255
Residuals	0.293	0.558	0.278	0.729

The table presents the variance analysis of the estimation reported in Table 3.7 column (5) and of the same regression when individual fixed effects (F.E.) are considered in the first step estimation, which is reported in Table 3.24 column (5) in Appendix 3.F. For the meaning of the figures reported, see the footnote in Table 3.5.

for some positions who publish more individually but exert a negative externality on other department academics. In general, we observe that less productive positions exert larger externalities (see the details of position externalities in Table 3.23 in Appendix 3.F).

The variables that really drive the differences in department externalities belong to the group we label as department 'research strategy' even if, as for size, their role somewhat reduces when individual fixed effects are controlled for in the first step. The first variable relates to the heterogeneity of academics in terms of publication record. The larger it is, the lower the department effects. More homogeneous departments, those where people have similar publication records, are those where local externalities are the strongest. Moreover, heterogeneity has the largest explanatory power of the department fixed effect since its effect presents both the largest standard error and the largest correlation with the dependent variable. See Table 3.8 for the average publication quality and Tables 3.25, 3.26 and 3.27 in Appendix 3.F for the probability to publish, the publication quantity and the publication top quality, respectively. When individual fixed effects are controlled for, increasing heterogeneity (that is decreasing homogeneity) by 25% (which corresponds to around half a standard error at the median) induces a decrease of the publication average top quality by 28%, of the publication quality by 11%, of the number of publications by 19% and of the probability to publish by 16%.[26] We are not aware of any similar finding in the literature.

Importantly, this positive effect of homogeneity within the department does not prevent star academics, who are in general an important source of heterogeneity, to also exert a positive externality. Having academics within the top 1% of academics in France is positive for all other academics belonging to the department. The effect is the largest for the publication top quality, then around three times smaller for publication quality, and then even smaller for quantity and the probability to publish. The explanatory power of this variable is pretty high also, much larger than for the department composition variables (size, age, women, position) even if around twice smaller than for heterogeneity. Attracting a star in a department is difficult since stars represent only 1% of the sample, by definition. However the return, on top of the individual effect captured in the first step, would be large according to our estimations. For a department of average size (around 30 academics), one more star means increasing the share of stars by around 0.03, which increases, when individual fixed effects are controlled for, the average publication top quality by 9.2%, the quality by 3.6% and the quantity by 7.2%.[27] These figures are of the same magnitude

26. $1.25^{-1.485}-1$, $1.25^{-0.504}-1$, $1.25^{-0.930}-1$, and $1.25^{-0.779}-1$, respectively, from the OLS estimates in Table 3.24 in Appendix 3.F.

27. $e^{2.925\times0.03}-1$, $e^{1.177\times0.03}-1$, and $e^{2.313\times0.03}-1$, respectively, from the OLS estimate in Table 3.24 in Appendix 3.F.

that the impact of dead stars on the publication record of their co-authors estimated by Azoulay et al. (2010), which is between 5 and 16%. An important difference here is that the effect regards the whole department and not only the star's co-authors.

The last department characteristic we study relates to its diversity in terms of research fields. This is the second characteristics, after heterogeneity, that has the highest explanatory power. It has a significant positive effect on all the publication dimensions, with the largest impact on the publication average top quality. Increasing diversity by 50% which is again close to one half a standard error at the median increase the average publication top quality by 36.3%, the quality by 10.4%, the quantity by 24.0% and the probability to publish by 17.3%.[28] Academics in departments with a share of publications similar in all fields do benefit from a positive externality from this variety of research fields they face.

3.7. Robustness checks: Teaching, spatial scale, and reverse causality

Even if we already mentioned some robustness of our results, we investigate now some further possible estimation issues.

3.7.1. Role of the teaching load

Research and teaching are often argued to be substitute activities for academics. Controlling for position fixed effects and position externalities is a first way to control for the differences in compulsory numbers of teaching hours by academics in France. Now, within

28. $1.50^{0.763} - 1$, $1.50^{0.244} - 1$, $1.50^{0.530} - 1$, and $1.50^{0.393} - 1$, respectively, from the OLS estimates in Table 3.24 in Appendix 3.F.

positions and even controlling for department size, departments possibly differ in terms of number of students involved in their programs. This can put more or less pressure on department academics and time for research. Alternatively, a larger share of graduate students for instance may exert a positive externality to the academics' publication activity. Therefore, testing which of the two effects dominates is interesting *per se*, on top of checking whether teaching is a possible missing variable that could bias some of our results.

Unfortunately, the number of students is reported in our data set only since 1999 and is not available for all universities, which largely reduces the number of observations, the reason why we do not include such effects in our main set of estimations. Still, since 1999 and for some universities, we know not only the total number of students in economics degrees (which we measure per academic present in the department since the role of department size is assessed separately) but also the share of students at the undergraduate level (the first two years after high school), graduate level (third and fourth years) and postgraduate level (fifth year and PhD).

Columns (1), (3), (5) and (7) in Table 3.9 replicate the corresponding columns in Table 3.7 but on the sub-sample of observations for which data on students is available. All results are fully consistent with those described in section 3.6, no effect changes sign or significance even if precision is sometimes a bit lost. The main conclusion is reached when the overall teaching load as well as the distribution of students between levels are controlled for, in columns (2), (4), (6) and (8). Our results on the determinants of the academic publication activity are therefore robust to the consideration of their second main task: teaching. Moreover, students do not seem to exert any positive or negative role on the publication activity. The only significant effect regards the fact that a larger share of undergraduate levels students increases the probability to publish. Then the number of publications or their quality are not affected by the number and composition of students.

Table 3.9: Role of teaching

	Publishing		Quantity		Quality		Top quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size	0.117^a	0.115^a	0.170^a	0.132^a	0.056^b	0.056^c	0.212^a	0.154^c
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)	(0.09)
Women	-0.101	-0.105	0.054	0.135	0.452^a	0.461^a	1.086^b	1.002^b
	(0.21)	(0.21)	(0.19)	(0.18)	(0.17)	(0.17)	(0.48)	(0.48)
Age	0.026^a	0.028^a	0.017^b	0.019^b	0.033^a	0.033^a	0.076^a	0.084^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Diversity	0.520^a	0.463^a	0.432^a	0.412^a	0.336^a	0.327^a	1.055^a	1.035^a
	(0.07)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.11)	(0.11)
Research Access	-0.017	0.007	-0.008	0.004	-0.008	-0.004	0.006	0.023
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)
Non-USA Openness	1.131^b	1.088^b	0.431	0.477	1.593^a	1.579^a	4.752^a	4.939^a
	(0.45)	(0.46)	(0.40)	(0.40)	(0.33)	(0.33)	(0.98)	(0.99)
USA Openness	1.977^b	2.627^a	2.099^a	2.222^a	0.761	0.845	4.002^b	4.494^a
	(0.81)	(0.88)	(0.68)	(0.72)	(0.54)	(0.56)	(1.61)	(1.63)
Heterogeneity	-1.043^a	-1.009^a	-1.044^a	-1.039^a	-0.853^a	-0.849^a	-2.806^a	-2.785^a
	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.16)	(0.16)
Stars	2.801^b	2.682^b	1.642	1.427	3.493^a	3.502^a	13.508^a	12.731^a
	(1.31)	(1.24)	(1.39)	(1.36)	(1.00)	(1.02)	(2.92)	(2.91)
Teaching-load		-0.020		0.047		0.005		-0.094
		(0.05)		(0.04)		(0.03)		(0.08)
% Undergraduate		0.766^a		0.162		0.129		0.060
		(0.16)		(0.16)		(0.13)		(0.37)
% Postgraduate		0.335		-0.057		0.076		-0.487
		(0.22)		(0.20)		(0.16)		(0.46)
Positions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.707	0.728	0.779	0.791	0.756	0.758	0.808	0.810
Observations	349	349	349	349	349	349	349	349

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

3.7.2. Spatial scale

We argue above that identifying whether externalities spill over the departments boundaries is interesting and we comment the role of the research assess variable along this line. Now, in many empirical economic geography studies the spatial scale at which estimations are performed is mainly guided by data availability. It can correspond as well to rather small units as cities, as larger ones as regions or states. We think that for the publication activity, the department is the most relevant scale at which scientific interactions take place, the reason why we chose it for our main set of estimations. Now, and closer to what is usually done in economic geography, we can consider something slightly larger that correspond to the city. We use the employment areas, which are 341 spatial units that fully cover France and were specifically built by INSEE, the French National Institute of Statistics, to study the role of local labour markets. For many employment areas, there are either no universities or only one (for 38 universities): Considering department or employment area is therefore the same. Six employment areas host two departments, four employment areas host three departments, and three employment areas host four or more departments. Tables 3.10 and 3.11 that replicate Tables 3.4 and 3.7 at the employment area level show that again no effect changes sign or significance by comparison with the department level. Therefore, it is possible that agglomeration and peer effects in economics operate at a geographical level slightly larger than the department. Importantly, this robustness of our conclusions to the level of spatial aggregation allows us to propose an instrumentation strategy to assess the possible role of reverse causality, an issue to which we turn now.

3.7.3. Reverse causality

It is possible that academics choose where to locate considering some of the variables we explain, typically the number or the quality of publications. As argued in section 3.2.2, this would create a reverse causality issue in the second step estimation for almost all variables we introduce. For instance Combes, Duranton and Gobillon (2008) show for productivity

Table 3.10: Determinants of individual publications at the employment area level

	Publishing		Quantity		Quality		Top quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Women	-0.375[a]	-0.400[a]	-0.090[a]	-0.421[a]	0.042[a]	-0.323[a]	0.109[b]	-1.043[a]
	(0.02)	(0.01)	(0.01)	(0.04)	(0.01)	(0.04)	(0.04)	(0.10)
Age	-0.105[a]	-0.097[a]	-0.031[a]	-0.116[a]	-0.000	-0.092[a]	-0.014	-0.299[a]
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Age square	0.000[a]	0.000[a]	0.000[a]	0.001[a]	-0.000	0.000[a]	0.000	0.001[a]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number			-0.920[a]	-0.889[a]	0.295[a]	0.293[a]	1.061[a]	1.037[a]
			(0.01)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
Non-USA Openness			0.363[a]	0.336[a]	0.256[a]	0.247[a]	0.895[a]	0.861[a]
			(0.01)	(0.01)	(0.02)	(0.01)	(0.06)	(0.03)
USA Openness			0.339[a]	0.315[a]	0.417[a]	0.392[a]	1.266[a]	1.180[a]
			(0.02)	(0.01)	(0.02)	(0.01)	(0.07)	(0.04)
Star			0.505[a]	0.441[a]	0.895[a]	0.828[a]	2.567[a]	2.342[a]
			(0.03)	(0.02)	(0.04)	(0.02)	(0.13)	(0.06)
Diversity			-0.131[a]	-0.084[a]	0.002	-0.003	0.103[a]	0.088[a]
			(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
Specialisation				0.344[a]		0.001		-0.001
				(0.01)		(0.01)		(0.01)
Quantity					0.088[a]	0.079[a]	0.495[a]	0.463[a]
					(0.01)	(0.01)	(0.02)	(0.01)
Selection			-0.164[a]	1.928[a]	-0.509[a]	1.732[a]	-1.631[a]	5.399[a]
			(0.04)	(0.24)	(0.06)	(0.24)	(0.17)	(0.62)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Jel Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Emp. Area Time FE	No	Yes	No	Yes	No	Yes	No	Yes
R^2			0.284	0.384	0.328	0.371	0.430	0.475
Observations	782730	778104	39330	39330	39330	39330	39330	39330

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. 'Publishing': Logit model for the probability to have at least one publication. 'Quantity', 'Quality', and 'Top quality': Heckman two-step Generalized Tobit models. The first step consists in a Probit model for the probability to publish. The variable 'Selection' in the second step is the inverse of Mills' ratio from the Probit equation. Variables are defined in section 3.2.

in market activities that the impact of city size would be overestimated by around 20% when using OLS for that reason. Unfortunately gathering instruments at the department level for variables as diverse as those we consider seems to be an impossible task. Our instrumentation strategy is thus based on two tricks.

First, we propose to evaluate the presence of reverse causality at the employment area level, for which more data is available while OLS results are very close to those at the

Table 3.11: Determinants of employment area fixed effects

	Publishing		Quantity		Quality		Top quality	
	OLS (1)	FGLS (2)	OLS (3)	FGLS (4)	OLS (5)	FGLS (6)	OLS (7)	FGLS (8)
Size	0.051^a	0.051^a	0.137^a	0.129^a	0.033^b	0.040^a	0.050	0.083^b
	(0.018)	(0.018)	(0.015)	(0.015)	(0.015)	(0.015)	(0.040)	(0.039)
Women	0.606^a	0.601^a	0.834^a	0.755^a	0.645^a	0.842^a	2.066^a	2.723^a
	(0.146)	(0.147)	(0.126)	(0.147)	(0.123)	(0.144)	(0.325)	(0.380)
Age	0.045^a	0.045^a	0.041^a	0.031^a	0.043^a	0.039^a	0.123^a	0.106^a
	(0.006)	(0.006)	(0.005)	(0.007)	(0.005)	(0.007)	(0.012)	(0.017)
Diversity	0.631^a	0.631^a	0.473^a	0.404^a	0.312^a	0.347^a	0.942^a	0.999^a
	(0.036)	(0.036)	(0.031)	(0.051)	(0.030)	(0.050)	(0.079)	(0.132)
Research Access	−0.008	−0.008	0.024^a	0.031^a	0.011	0.017^b	0.046^b	0.055^b
	(0.010)	(0.010)	(0.008)	(0.009)	(0.008)	(0.008)	(0.022)	(0.022)
Non-USA Openness	1.796^a	1.797^a	2.040^a	1.949^a	2.148^a	1.966^a	6.954^a	6.502^a
	(0.333)	(0.333)	(0.288)	(0.333)	(0.280)	(0.326)	(0.741)	(0.861)
USA Openness	1.802^a	1.800^a	0.367	0.163	0.479	0.991^b	2.535^b	3.554^a
	(0.528)	(0.528)	(0.457)	(0.442)	(0.444)	(0.428)	(1.176)	(1.135)
Heterogeneity	-1.146^a	-1.145^a	-1.012^a	-0.915^a	-0.809^a	-0.850^a	-2.506^a	-2.567^a
	(0.057)	(0.057)	(0.049)	(0.116)	(0.048)	(0.115)	(0.126)	(0.302)
Stars	3.152^b	3.177^b	5.683^a	5.970^a	5.735^a	6.202^a	19.793^a	20.465^a
	(1.232)	(1.232)	(1.065)	(1.017)	(1.036)	(0.985)	(2.743)	(2.610)
Positions	yes	yes	yes	yes	yes	yes	yes	yes
Time Fixed Effect	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.80	0.80	0.87	0.93	0.75	0.86	0.80	0.89
Observations	767	767	767	767	767	767	767	767

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

departments level as we have just shown. For instance, it is pretty clear that the number of academics in economics is positively correlated to the overall population of the employment area, large cities hosting in general large universities. Now, the correlation between these two variables is far from perfect, which is an interesting property that should makes it easier to satisfy the exogeneity condition of the instruments. Indeed, not all cities have the same tradition as regards the presence of universities and some smaller cities may have larger universities. Moreover, history made some cities specialising in different fields and what we instrument is the size of the economics department, not the total university size.

The second trick consists in instrumenting variables one by one without any further control variable. Instrumenting all variables simultaneously would require lots of instruments and, most importantly, over-identification tests become doubtful when the number of instrumented variables and of instruments become too large. Moreover, if instruments are shown to be valid, instrumentation solves both reverse causality and missing variables issues, and therefore instrumental variables techniques should provide a consistent estimate for the effect of each variable even when introduced alone. Last, notice that our purpose here is not to obtain definite values for each of the effects we estimate but only to confirm that no major endogeneity issue biases our estimates, and change their sign or significance for instance.

In order to perform meaningful over-identification tests, we need many instruments, possibly different in their nature. Therefore, next to the employment area population in 1999 that is pretty obviously a determinant of many of the explanatory variables we consider as we have just argued for department size, we also use the share of engineers in local employment (still in 1999). The intuition is that hard science universities or Grandes Ecoles are often located in areas where high-tech industries, and thus engineer occupations, are over-represented. Then, given the French tradition of centralisation and following other studies on French data, we use a physical geography variable that consists in a peripherality index. It is the average distance of the employment area to all other employment areas (without any weight, by population or employment for instance, to reduce possible endogeneity).

Tables 3.12 to 3.15 present OLS and IV estimates for the role of some of the most important variables of the second step: size, diversity, heterogeneity and stars. To assess the quality of the instrumentation, we report the Shea partial R^2, the p-value of the over-identification test and the Cragg-Donald statistics that checks for the possible weakness of the instruments. Over-identification tests are passed and the instruments are not weak, in the sense that the lowest Cragg-Donald value is 12. The conclusion is that the impact of size, diversity, heterogeneity and stars on any component of research productivity is robust to instrumentation. If anything, instrumental variable estimates are of larger magnitude than OLS ones. Clearly, better assessing the role of reverse causality, which is almost never done in the literature on peer effects in academia contrary to those in other domains (see for instance Bramoullé et al. (2009) for a recent contribution), remains high on the research agenda but at least such results make us confident that the OLS estimates we present should not largely over-estimate the true impact of the variables.

Table 3.12: Instrumental variable estimates for the effect of size

	Publishing		Quantity		Quality		Top quality	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size	0.140^a	0.152^a	0.232^a	0.294^a	0.107^a	0.189^a	0.311^a	0.550^a
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.06)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shea p. R^2		0.416		0.416		0.416		0.416
J-stat p-value		0.969		0.382		0.261		0.399
Cragg-Donald		255.16		255.16		255.16		255.16
R^2	0.589	0.588	0.481	0.472	0.501	0.481	0.495	0.475
Observations	735	735	735	735	735	735	735	735

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. Instruments for size are the logarithm of the employment area population and the share of engineers in employment in 1990. The first step of the IV two stages least squares is in the first column of Table 3.30 in Appendix 3.H.

3.8. Conclusion

Location matters for the publication performance of academics. A careful variance analysis of individual publication determinants shows that the explanatory power of de-

Table 3.13: Instrumental variable estimates for the effect of diversity

	Publishing		Quantity		Quality		Top quality	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Diversity	0.906^a	1.227^a	0.876^a	2.137^a	0.540^a	1.453^a	1.635^a	4.197^a
	(0.05)	(0.14)	(0.05)	(0.29)	(0.04)	(0.20)	(0.12)	(0.57)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shea p. R^2		0.079		0.045		0.043		0.043
J-stat p-value		0.247		0.006		0.117		0.080
Cragg-Donald		30.85		16.71		16.19		16.19
R^2	0.735	0.713	0.622	0.086	0.601	0.215	0.606	0.250
Observations	735	735	735	735	735	735	735	735

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. Instruments for diversity are the logarithm of population in 1990 and peripherality for publishing, the logarithme of population in 1999 and the share of engineers for quantity and the logarithm of the population and the share of engineers in employment in 1990 quality and top quality. The first steps of the IV two stages least squares are in columns (2), (3) and (4) of Table 3.30 in Appendix 3.H.

Table 3.14: Instrumental variable estimates for the effect of heterogeneity

	Publishing		Quantity		Quality		Top quality	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Heterogeneity	-1.209^a	-2.021^a	-1.043^a	-2.903^a	-0.788^a	-2.325^a	-2.515^a	-6.573^a
	(0.08)	(0.44)	(0.07)	(0.61)	(0.05)	(0.43)	(0.15)	(1.27)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shea p. R^2		0.032		0.032		0.032		0.032
J-stat p-value		0.207		0.214		0.673		0.566
Cragg-Donald		11.80		11.80		11.80		11.80
R^2	0.703	0.639	0.530	-0.002	0.598	0.096	0.617	0.210
Observations	735	735	735	735	735	735	735	735

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. Only heterogeneity has been instrumented at the employment area level with peripherality and share of engineers. First step of the IV two stages least squares is in column (5) of Table 3.30 in Appendix 3.H.

partments effects represents at least half of individual effects' one. When explaining department performance, selection and local effects have similar explanatory power. As argued by Waldinger (2012), this corresponds better to what many academics have in mind if one considers the time they spend assessing the relative qualities of departments. This is in sharp contrast with previous findings from the literature that concludes to the presence of

Table 3.15: Instrumental variable estimates for the effect of stars

	Publishing		Quantity		Quality		Top quality	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stars	8.511^a	22.238^a	11.151^a	32.969^a	9.618^a	26.081^a	31.554^a	73.877^a
	(1.33)	(4.35)	(1.76)	(4.93)	(1.22)	(3.44)	(3.48)	(9.80)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shea p. R^2		0.128		0.128		0.128		0.128
J-stat p-value		0.093		0.010		0.461		0.273
Cragg-Donald		52.82		52.82		52.82		52.82
R^2	0.576	0.540	0.401	0.255	0.505	0.390	0.510	0.422
Observations	735	735	735	735	735	735	735	735

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. Only the share of stars has been instrumented at the employment area level with peripherality and share of engineers. First step of the IV two stages least squares is in column (6) of Table 3.30 in Appendix 3.H.

small, if not totally absent, local effects. We attribute this difference in conclusions to the fact that we have access to an exhaustive data set of all academic economists in France whom we can follow over time and across locations even when they do not publish, which present the further advantage of not biasing the computation of department characteristics. Moreover, we also have access to more individual variables, some of them that vary over time, which is in general not available. We also separately study the determinants of the probability to publish, the number of publications and their average quality when only a quality adjusted number of publications is generally considered in the literature.

Due to numerous possible sources of missing variable and reverse causality issues at stake when estimating agglomeration and peer effects, we do not claim to provide a final assessment of the role of department characteristics on individual performance. The possibility to merge together bibliometric and administrative sources as we do here will certainly be extended in the future (to longer periods, other fields and countries) and this should allow the researchers to find even more sources of exogenous variations to properly assess the role of endogenous and exogenous individual and department characteristics. Putting more structure to the underlying models of network formation and of agglomeration and peer effects, which are only implicit here and therefore treated as black box, should also certainly help to improve the estimated specifications. Ultimately important results should be obtained for a better design of research policies.

APPENDIX

3.A. Top 50 journals

Table 3.16: Top 50 journals

Journal	Rank	Quality	Top quality
quarterly journal of economics	1	100.0	100.0
american economic review	2	98.1	94.4
journal of political economy	3	96.2	89.1
econometrica	4	95.7	87.7
review of economic studies	5	81.0	53.1
journal of financial economics	6	80.6	52.4
journal of monetary economics	7	75.8	43.6
review of economics and statistics	8	74.1	40.7
journal of economic theory	9	72.8	38.5
journal of finance	10	72.2	37.6
journal of econometrics	11	68.6	32.3
economic journal	12	64.5	26.8
rand journal of economics	13	63.7	25.8
journal of public economics	14	62.0	23.9
journal of international economics	15	61.5	23.3
journal of the european economic association	16	57.0	18.5
european economic review	17	55.2	16.8
journal of labor economics	18	55.1	16.7
international economic review	19	54.7	16.4
games and economic behavior	20	54.1	15.8
review of financial studies	21	49.1	11.8
journal of business and economic statistics	22	48.1	11.1
journal of health economics	23	43.9	8.5
journal of development economics	24	42.7	7.8
journal of human resources	25	42.2	7.5
journal of money, credit, and banking	26	41.9	7.3
journal of law and economics	27	40.7	6.8
journal of accounting and economics	28	40.5	6.6
journal of urban economics	29	40.0	6.4
journal of environmental economics and management	30	37.6	5.3
journal of economic growth	31	37.4	5.2
journal of economic dynamics and control	32	36.1	4.7
journal of economic behavior and organization	33	35.8	4.6
world development	34	35.8	4.6
review of economic dynamics	35	35.3	4.4
journal of applied econometrics	36	35.0	4.3
economic theory	37	34.0	3.9
econometric theory	38	33.7	3.8
journal of law, economics, and organization	39	32.1	3.3
health economics	40	31.5	3.1
american journal of agricultural economics	41	31.4	3.1
journal of industrial economics	42	31.1	3.0
international journal of industrial organization	43	31.0	3.0
journal of economic history	44	31.0	3.0
journal of economic perspectives	45	30.5	2.8
economics letters	46	30.4	2.8
journal of risk and uncertainty	47	30.0	2.7
scandinavian journal of economics	48	30.0	2.7
journal of financial and quantitative analysis	49	29.7	2.6
ecological economics	50	29.5	2.6

3.B. Determinants of individual publications with individual fixed effects

Table 3.17: Determinants of individual publications with individual fixed effects

	Publishing		Quantity		Quality		Top quality	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Age	0.129^a	0.108^a	0.021^a	-0.056^c	0.035^a	-0.021	0.114^a	-0.050
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.07)
Age square	-0.001^a	-0.001^a	-0.000^b	0.000^a	-0.000^a	0.000	-0.001^a	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number			-0.901^a	-0.856^a	0.253^a	0.252^a	0.877^a	0.882^a
			(0.02)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
Non-USA Openness			0.190^a	0.174^a	0.056^a	0.063^a	0.241^a	0.246^a
			(0.02)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
USA Openness			0.215^a	0.196^a	0.195^a	0.181^a	0.522^a	0.488^a
			(0.03)	(0.02)	(0.02)	(0.02)	(0.07)	(0.04)
Star			0.310^a	0.287^a	0.587^a	0.583^a	1.574^a	1.563^a
			(0.03)	(0.03)	(0.05)	(0.02)	(0.12)	(0.06)
Diversity			-0.146^a	-0.092^a	0.004	0.004	0.047	0.052^a
			(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
Specialisation				0.346^a		0.008^c		0.035^a
				(0.00)		(0.00)		(0.01)
Quantity					0.059^a	0.056^a	0.356^a	0.346^a
					(0.01)	(0.00)	(0.02)	(0.01)
Selection			-0.085^c	2.310^a	0.043	1.552^a	0.083	4.550^a
			(0.05)	(0.38)	(0.05)	(0.35)	(0.13)	(0.89)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Jel Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department Time FE	No	Yes	No	Yes	No	Yes	No	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2			0.506	0.599	0.630	0.660	0.709	0.731
Observations	427662	425178	38984	38984	38984	38984	38984	38984

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. 'Publishing': Logit model for the probability to have at least one publication. 'Quantity', 'Quality', and 'Top quality': Heckman two-step Generalized Tobit models. The first step consists in a Probit model for the probability to publish. The variable 'Selection' in the second step is the inverse of Mills' ratio from the Probit equation. Variables are defined in section 3.2.

3.C. Variance analysis of individual publication quality

Table 3.18: Variance analysis of the individual average publication quality

	Without individual fixed effects			With individual fixed effects		
	Stand. error	Correlation	Sorting	Stand. error	Correlation	Sorting
Explained: quality	0.452	1.000		0.452	1.000	
Individual effects	0.482	0.424	0.188	0.503	0.673	0.089
Individual fixed effect	-	-	-	0.330	0.696	0.076
Gender	0.068	0.050	0.006	-	-	-
Age	0.285	0.132	-0.052	0.097	0.133	-0.029
Position fixed effect	0.273	0.157	0.199	0.138	0.082	0.007
Authors number	0.060	0.215	0.059	0.054	0.215	0.045
Non-USA openness	0.048	0.277	0.133	0.012	0.277	0.070
USA openness	0.055	0.317	0.173	0.027	0.317	0.106
Star	0.072	0.338	0.205	0.054	0.338	0.124
Individual diversity	0.002	-0.141	-0.187	0.001	0.141	0.156
Jel fixed effect	0.260	0.204	0.066	0.247	0.174	0.034
Department effects	0.237	0.367	1.000	0.156	0.242	1.000
Department fixed effect	0.238	0.366	1.000	0.157	0.240	1.000
Specialisation	0.003	0.039	-0.145	0.004	0.039	-0.100
Quantity	0.034	0.134	0.088	0.025	0.134	0.034
Selection	0.390	-0.305	-0.425	0.297	-0.280	-0.316
Residuals	0.352	0.779	0	0.265	0.586	0

The table presents the variance analysis of the estimation reported in Table 3.4 column (6) and of the same regression including also individual fixed effects that is reported in Table 3.17 column (6) in 3.B. For the meaning of the figures reported, see the footnote in Table 3.5.

Table 3.19: Variance analysis of the individual publication top quality

	Without individual fixed effects			With individual fixed effects		
	Stand. error	Correlation	Sorting	Stand. error	Correlation	Sorting
Explained: top quality	1.303	1.000		1.303	1.000	
Individual effects	1.610	0.484	0.202	1.517	0.724	0.139
Individual fixed effect	-	-	-	0.950	0.736	0.148
Gender	0.235	0.070	0.005	-	-	-
Age	0.943	0.142	-0.047	0.357	0.144	-0.032
Position fixed effect	0.926	0.198	0.205	0.432	0.135	0.037
Authors number	0.213	0.236	0.062	0.190	0.236	0.032
Non-USA openness	0.169	0.335	0.140	0.048	0.335	0.082
USA openness	0.166	0.360	0.182	0.072	0.360	0.115
Star	0.200	0.375	0.218	0.145	0.375	0.131
Individual diversity	0.024	0.179	0.192	0.017	0.179	0.161
Jel fixed effect	0.874	0.228	0.073	0.724	0.201	0.036
Department effects	0.758	0.400	1.000	0.416	0.267	0.999
Department fixed effect	0.760	0.398	1.000	0.417	0.263	1.000
Specialisation	0.012	0.071	-0.144	0.017	0.071	-0.117
Quantity	0.202	0.214	0.096	0.158	0.214	0.041
Selection	1.313	-0.364	-0.445	0.869	-0.337	-0.335
Residuals	0.924	0.709	0	0.678	0.520	0

The table presents the variance analysis of the estimation reported in Table 3.4 column (8) and of the same regression including also individual fixed effects that is reported in Table 3.17 column (8) in 3.B. For the meaning of the figures reported, see the footnote in Table 3.5.

3.D. Determinants of individual publications with the details of position effects

Table 3.20: Determinants of individual publication with the details of position effects

	Publishing		Quantity		Quality		Top quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Women	-0.368^a	-0.373^a	-0.090^a	-0.376^a	0.043^a	-0.297^a	0.102^b	-1.024^a
	(0.02)	(0.01)	(0.01)	(0.04)	(0.01)	(0.04)	(0.04)	(0.09)
Age	-0.104^a	-0.090^a	-0.030^a	-0.107^a	0.002	-0.084^a	-0.010	-0.292^a
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Age square	0.000^a	0.000^a	0.000^a	0.001^a	-0.000	0.000^a	0.000	0.001^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No 05	-0.974^a	-0.933^a	0.104^a	-0.671^a	0.221^a	-0.597^a	0.742^a	-1.958^a
	(0.05)	(0.02)	(0.02)	(0.09)	(0.03)	(0.09)	(0.09)	(0.23)
No 37	-1.534^a	-1.411^a	-0.190^a	-1.247^a	0.110	-1.168^a	0.422^c	-3.684^a
	(0.14)	(0.09)	(0.07)	(0.13)	(0.09)	(0.13)	(0.24)	(0.35)
A-CR	-0.530^a	-0.432^a	0.036	-0.275^a	0.189^a	-0.191^a	0.644^a	-0.626^a
	(0.10)	(0.06)	(0.04)	(0.05)	(0.06)	(0.05)	(0.17)	(0.14)
A-MCF	-2.147^a	-2.232^a	0.205^a	-1.227^a	0.458^a	-1.187^a	1.351^a	-4.018^a
	(0.16)	(0.10)	(0.07)	(0.18)	(0.06)	(0.18)	(0.17)	(0.48)
A-PR	0.727^a	1.079^a	0.232^b	1.318^a	0.359^a	1.076^a	1.220^a	3.709^a
	(0.26)	(0.06)	(0.12)	(0.12)	(0.07)	(0.12)	(0.20)	(0.32)
Other	-0.987^a	-0.971^a	0.328^a	-0.514^a	0.412^b	-0.463^a	1.062^c	-1.866^a
	(0.28)	(0.16)	(0.10)	(0.15)	(0.20)	(0.15)	(0.62)	(0.39)
CR cnrs	0.737^a	0.556^a	0.121^a	0.560^a	0.078^a	0.564^a	0.210^a	1.815^a
	(0.04)	(0.03)	(0.02)	(0.06)	(0.03)	(0.06)	(0.08)	(0.15)
CR inra	0.763^a	0.683^a	0.060^b	0.561^a	0.060^c	0.492^a	0.241^b	1.774^a
	(0.06)	(0.04)	(0.03)	(0.07)	(0.04)	(0.07)	(0.10)	(0.19)
DE ehess	2.625^a	2.002^a	0.155^b	1.724^a	-0.068	1.640^a	0.334	5.864^a
	(0.09)	(0.07)	(0.07)	(0.20)	(0.09)	(0.20)	(0.25)	(0.52)
DR cnrs	1.623^a	1.384^a	0.230^a	1.288^a	0.006	1.175^a	-0.022	3.777^a
	(0.05)	(0.03)	(0.03)	(0.13)	(0.04)	(0.13)	(0.11)	(0.35)
DR inra	1.314^a	1.175^a	0.140^a	1.014^a	-0.114^a	0.787^a	-0.357^a	2.735^a
	(0.05)	(0.04)	(0.03)	(0.12)	(0.04)	(0.12)	(0.11)	(0.31)
PR	1.143^a	1.062^a	0.176^a	1.019^a	-0.094^a	0.845^a	-0.152^b	2.933^a
	(0.03)	(0.01)	(0.02)	(0.10)	(0.02)	(0.10)	(0.07)	(0.27)
Insee	1.744^a	1.287^a	0.209^a	1.102^a	0.218^b	1.150^a	0.779^a	3.844^a
	(0.14)	(0.07)	(0.05)	(0.13)	(0.09)	(0.14)	(0.20)	(0.36)
IPC	1.998^a	1.464^a	0.342^a	1.358^a	0.124^b	1.232^a	0.577^a	4.131^a
	(0.11)	(0.08)	(0.06)	(0.14)	(0.06)	(0.15)	(0.18)	(0.38)
Authors number			-0.917^a	-0.868^a	0.292^a	0.279^a	1.050^a	0.989^a
			(0.01)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
Non-USA Openness			0.362^a	0.319^a	0.253^a	0.242^a	0.889^a	0.854^a
			(0.01)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)
USA Openness			0.338^a	0.322^a	0.408^a	0.377^a	1.243^a	1.129^a
			(0.02)	(0.01)	(0.03)	(0.01)	(0.07)	(0.04)
Star			0.492^a	0.413^a	0.877^a	0.772^a	2.504^a	2.156^a
			(0.03)	(0.02)	(0.04)	(0.02)	(0.12)	(0.06)
Diversity			-0.131^a	-0.060^a	-0.002	-0.007	0.087^a	0.074^a
			(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
Selection			-0.180^a	1.861^a	-0.565^a	1.662^a	-1.743^a	5.600^a
			(0.04)	(0.23)	(0.05)	(0.23)	(0.14)	(0.61)
Specialisation				0.368^a		0.007		0.024^c
				(0.00)		(0.01)		(0.01)
Quantity					0.087^a	0.075^a	0.491^a	0.442^a
					(0.01)	(0.01)	(0.02)	(0.01)
Jel Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department Time FE	No	Yes	No	Yes	No	Yes	No	Yes
R^2			0.284	0.425	0.330	0.398	0.433	0.501
Observations	771426	760122	38984	38984	38984	38984	38984	38984

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. 'Publishing': Logit model for the probability to have at least one publication. 'Quantity', 'Quality', and 'Top quality': Heckman two-step Generalized Tobit models. The first step consists in a Probit model for the probability to publish. The variable 'Selection' in the second step is the inverse of Mills' ratio from the Probit equation. Variables are defined in section 3.2. A-CR: other research fellow; A-MCF: non-university assistant professor; A-PR: non-university professor; CR cnrs: Cnrs research fellow; CR inra: Inra research fellow; DE ehess: Ehess research professor; DR cnrs: Cnrs research professor; DR inra: Inra research professor; PR: university professor; Insee: national statistical institute engineers; IPC: Ponts-et-Chaussées engineers. Reference: assistant professor.

3.E. Variance analysis at the department level

Table 3.21: Variance analysis of the quantity published at the department level

	Without individual fixed effects			With individual fixed effects		
	Stand. error	Correlation	Sorting	Stand. error	Correlation	Sorting
Explained: quantity	0.373	1.000		0.373	1.000	
Individual effects	0.432	0.413	0.242	0.527	0.496	0.047
Individual fixed effect	-	-	-	0.316	0.359	-0.139
Gender	0.075	0.112	0.059	-	-	-
Age	0.278	-0.065	-0.074	0.106	-0.056	-0.012
Position fixed effect	0.347	0.194	0.222	0.240	0.159	0.111
Authors number	0.211	0.486	0.018	0.208	0.486	0.037
Non-USA openness	0.052	0.074	0.096	0.028	0.074	0.081
USA openness	0.033	0.109	0.252	0.020	0.109	0.204
Star	0.031	0.242	0.296	0.022	0.242	0.253
Individual diversity	0.018	0.118	-0.353	0.027	0.118	-0.347
Jel fixed effect	0.174	0.007	0.135	0.236	0.004	0.142
Department effects	0.500	0.555	0.946	0.482	0.364	0.948
Department fixed effect	0.617	0.446	1.000	0.591	0.293	1.000
Specialisation	0.217	0.011	-0.666	0.204	0.011	-0.656
Selection	0.591	-0.140	-0.696	0.575	-0.111	-0.649
Residuals	0.001	-0.062	0	0.001	-0.037	0

The table presents the variance analysis of the estimation reported in Table 3.4 column (4) and of the same regression including also individual fixed effects that is reported in Table 3.17 column (4) in 3.B, once they are averaged by department. For the meaning of the figures reported, see the footnote in Table 3.5.

Table 3.22: Variance analysis of the publication top quality at the department level

	Without individual fixed effects			With individual fixed effects		
	Stand. error	Correlation	Sorting	Stand. error	Correlation	Sorting
Explained: top quality	1.282	1.000		1.282	1.000	
Individual effects	1.652	0.583	0.250	1.619	0.784	0.115
Individual fixed effect	-	-	-	1.096	0.763	0.096
Gender	0.204	0.070	0.006	-	-	-
Age	0.883	0.090	-0.113	0.327	0.080	-0.038
Position fixed effect	1.020	0.369	0.274	0.372	0.224	-0.005
Authors number	0.241	0.348	0.074	0.215	0.348	0.073
Non-USA openness	0.139	0.453	0.193	0.040	0.453	0.129
USA openness	0.115	0.525	0.343	0.050	0.525	0.256
Star	0.162	0.567	0.384	0.117	0.567	0.230
Individual diversity	0.022	0.247	0.390	0.015	0.247	0.347
Jel fixed effect	0.721	0.263	0.108	0.604	0.222	0.048
Department effects	1.615	0.661	1.000	1.033	0.449	1.000
Department fixed effect	1.622	0.660	1.000	1.041	0.448	1.000
Specialisation	0.014	-0.124	-0.461	0.020	-0.124	-0.409
Quantity	0.165	0.215	0.140	0.129	0.215	-0.003
Selection	1.778	-0.442	-0.678	1.132	-0.423	-0.568
Residuals	0.001	-0.030	0	0.001	-0.001	0

The table presents the variance analysis of the estimation reported in Table 3.4 column (8) and of the same regression including also individual fixed effects that is reported in Table 3.17 column (8) in 3.B, once they are averaged by department. For the meaning of the figures reported, see the footnote in Table 3.5.

3.F. Second step analysis

Table 3.23: Determinants of department fixed effects with detailed position effects

	Publishing		Quantity		Quality		Top quality	
	OLS	FGLS	OLS	FGLS	OLS	FGLS	OLS	FGLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size	0.100^{a}	0.101^{a}	0.198^{a}	0.174^{a}	0.067^{a}	0.078^{a}	0.172^{a}	0.192^{a}
	(0.014)	(0.014)	(0.012)	(0.012)	(0.013)	(0.013)	(0.033)	(0.034)
Women	0.230^{b}	0.229^{b}	0.490^{a}	0.355^{a}	0.334^{a}	0.448^{a}	1.173^{a}	1.558^{a}
	(0.102)	(0.104)	(0.090)	(0.109)	(0.095)	(0.114)	(0.247)	(0.297)
Age	0.034^{a}	0.033^{a}	0.029^{a}	0.021^{a}	0.025^{a}	0.022^{a}	0.081^{a}	0.072^{a}
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.010)	(0.013)
Diversity	0.662^{a}	0.662^{a}	0.448^{a}	0.367^{a}	0.323^{a}	0.329^{a}	1.037^{a}	1.019^{a}
	(0.026)	(0.026)	(0.023)	(0.046)	(0.024)	(0.047)	(0.063)	(0.123)
Research Access	-0.017^{b}	-0.017^{b}	-0.011	-0.005	0.015^{b}	0.013^{c}	0.042^{b}	0.031
	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.020)	(0.020)
Non-USA Openness	1.287^{a}	1.290^{a}	1.451^{a}	1.262^{a}	1.435^{a}	1.347^{a}	4.941^{a}	4.680^{a}
	(0.202)	(0.202)	(0.179)	(0.220)	(0.188)	(0.229)	(0.491)	(0.597)
USA Openness	0.858^{a}	0.853^{a}	0.210	0.088	1.023^{a}	1.084^{a}	3.522^{a}	3.563^{a}
	(0.276)	(0.276)	(0.245)	(0.228)	(0.257)	(0.241)	(0.669)	(0.628)
Heterogeneity	-1.147^{a}	-1.145^{a}	-1.082^{a}	-0.900^{a}	-0.752^{a}	-0.757^{a}	-2.539^{a}	-2.493^{a}
	(0.039)	(0.039)	(0.035)	(0.105)	(0.037)	(0.107)	(0.095)	(0.280)
Stars	0.828^{b}	0.832^{b}	1.549^{a}	1.387^{a}	2.527^{a}	2.538^{a}	7.691^{a}	7.759^{a}
	(0.375)	(0.375)	(0.332)	(0.328)	(0.349)	(0.345)	(0.909)	(0.899)
A-CR	0.268^{c}	0.241	0.346^{a}	0.212^{c}	0.045	0.098	0.147	0.304
	(0.145)	(0.171)	(0.129)	(0.120)	(0.135)	(0.127)	(0.352)	(0.330)
A-MCF	1.259^{a}	1.261^{a}	0.912^{a}	0.809^{a}	1.090^{a}	1.230^{a}	3.607^{a}	3.977^{a}
	(0.142)	(0.143)	(0.126)	(0.178)	(0.132)	(0.184)	(0.345)	(0.480)
A-PR	-1.105^{a}	-1.112^{a}	-1.400^{a}	-1.330^{a}	-0.487^{b}	-0.506^{b}	-1.926^{a}	-1.922^{a}
	(0.238)	(0.278)	(0.211)	(0.204)	(0.222)	(0.215)	(0.578)	(0.561)
Other	3.769^{a}	3.754^{a}	4.229^{a}	3.184^{a}	1.790	1.569	4.993	4.520
	(1.317)	(1.329)	(1.168)	(1.018)	(1.225)	(1.083)	(3.193)	(2.816)
CR cnrs	0.105	0.099	0.010	0.024	0.147	0.100	0.235	0.311
	(0.157)	(0.162)	(0.139)	(0.140)	(0.146)	(0.147)	(0.380)	(0.383)
CR inra	-0.415^{a}	-0.440^{a}	-0.192^{c}	-0.163	-0.175	-0.155	-0.564^{c}	-0.534^{c}
	(0.126)	(0.141)	(0.111)	(0.115)	(0.117)	(0.121)	(0.305)	(0.315)
DE ehess	-0.103	-0.185	-0.051	-0.044	-1.341^{a}	-1.355^{a}	-3.599^{a}	-3.663^{a}
	(0.427)	(0.447)	(0.378)	(0.333)	(0.397)	(0.353)	(1.034)	(0.919)
DR cnrs	-0.562^{a}	-0.537^{a}	0.027	0.068	-0.165	-0.046	-0.124	0.178
	(0.135)	(0.142)	(0.119)	(0.117)	(0.125)	(0.123)	(0.326)	(0.320)
DR inra	-0.251^{b}	-0.263^{c}	-0.162	-0.088	0.188^{c}	0.214^{c}	0.197	0.346
	(0.117)	(0.135)	(0.104)	(0.108)	(0.109)	(0.113)	(0.283)	(0.294)
PR	-0.458^{a}	-0.443^{a}	0.258^{a}	0.281^{a}	0.061	0.106	0.428^{c}	0.578^{b}
	(0.102)	(0.107)	(0.090)	(0.093)	(0.095)	(0.098)	(0.247)	(0.255)
Insee	-0.561^{c}	-0.522	-0.081	0.003	0.260	0.307	0.926	1.135^{c}
	(0.289)	(0.321)	(0.257)	(0.210)	(0.269)	(0.224)	(0.701)	(0.583)
IPC	-0.913^{a}	-0.945^{a}	-0.395^{b}	-0.317	-1.286^{a}	-1.183^{a}	-3.847^{a}	-3.515^{a}
	(0.222)	(0.324)	(0.197)	(0.222)	(0.207)	(0.232)	(0.539)	(0.604)
No-05	0.343^{a}	0.326^{a}	0.176^{b}	0.173^{b}	-0.171^{b}	-0.184^{b}	-0.004	-0.181
	(0.091)	(0.096)	(0.081)	(0.077)	(0.085)	(0.082)	(0.221)	(0.212)
No-37	-0.487	-0.494	-0.081	-0.018	-0.394	-0.243	-1.762	-1.358
	(0.529)	(0.531)	(0.469)	(0.432)	(0.492)	(0.458)	(1.283)	(1.192)
Time Fixed Effect	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.82	0.82	0.86	0.93	0.75	0.85	0.82	0.90
Observations	1209	1209	1209	1209	1209	1209	1209	1209

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. A-CR: other research fellow; A-MCF: non-university assistant professor; A-PR: non-university professor; CR cnrs: Cnrs research fellow; CR inra: Inra research fellow; DE ehess: Ehess research professor; DR cnrs: Cnrs research professor; DR inra: Inra research professor; PR: university professor; Insee: national statistical institute engineers; IPC: Ponts-et-Chaussées engineers. Reference: assistant professor.

Table 3.24: Determinants of department fixed effects with individual fixed effects

	Publishing		Quantity		Quality		Top quality	
	OLS (1)	FGLS (2)	OLS (3)	FGLS (4)	OLS (5)	FGLS (6)	OLS (7)	FGLS (8)
Size	−0.010	0.003	0.161^a	0.114^a	0.020	0.014	0.057^c	0.054
	(0.021)	(0.025)	(0.013)	(0.016)	(0.012)	(0.014)	(0.030)	(0.035)
Women	0.653^a	0.283	0.867^a	0.572^a	0.367^a	0.428^a	0.412^c	0.872^b
	(0.154)	(0.195)	(0.099)	(0.151)	(0.090)	(0.137)	(0.224)	(0.346)
Age	0.054^a	0.045^a	0.036^a	0.034^a	0.020^a	0.021^a	0.058^a	0.054^a
	(0.006)	(0.007)	(0.004)	(0.008)	(0.004)	(0.007)	(0.009)	(0.018)
Diversity	0.393^a	0.511^a	0.530^a	0.467^a	0.244^a	0.269^a	0.763^a	0.789^a
	(0.039)	(0.041)	(0.025)	(0.073)	(0.023)	(0.066)	(0.057)	(0.169)
Research Access	-0.062^a	-0.073^a	-0.077^a	-0.058^a	-0.016^b	-0.018^c	0.007	−0.019
	(0.012)	(0.018)	(0.008)	(0.011)	(0.007)	(0.010)	(0.018)	(0.026)
Non-USA Openness	−0.025	−0.357	1.496^a	1.160^a	0.510^a	0.718^a	1.906^a	2.238^a
	(0.306)	(0.310)	(0.197)	(0.287)	(0.179)	(0.260)	(0.446)	(0.657)
USA Openness	1.457^a	0.636	0.843^a	0.617^b	0.955^a	0.802^a	3.151^a	2.285^a
	(0.417)	(0.400)	(0.269)	(0.292)	(0.243)	(0.265)	(0.608)	(0.661)
Heterogeneity	-0.779^a	-0.853^a	-0.930^a	-0.814^a	-0.504^a	-0.525^a	-1.485^a	-1.514^a
	(0.059)	(0.060)	(0.038)	(0.127)	(0.035)	(0.115)	(0.087)	(0.294)
Stars	-1.978^a	−0.881	2.313^a	1.838^a	1.177^a	1.262^a	2.925^a	3.436^a
	(0.567)	(0.555)	(0.365)	(0.395)	(0.331)	(0.358)	(0.826)	(0.895)
A-CR	-0.656^a	−0.395	0.158	0.392^c	0.150	0.140	0.089	0.155
	(0.219)	(0.386)	(0.141)	(0.220)	(0.128)	(0.199)	(0.320)	(0.503)
A-MCF	−0.172	0.377	0.339^b	0.395^b	0.444^a	0.528^a	1.230^a	1.269^a
	(0.215)	(0.269)	(0.139)	(0.201)	(0.126)	(0.182)	(0.313)	(0.459)
A-PR	0.323	0.556	-1.397^a	-1.185^a	-1.009^a	-0.949^a	-2.199^a	-1.943^b
	(0.360)	(0.382)	(0.232)	(0.341)	(0.210)	(0.310)	(0.526)	(0.782)
Other	1.878	2.238	2.465^c	1.734	0.635	0.824	2.012	1.371
	(1.992)	(1.961)	(1.283)	(1.191)	(1.163)	(1.078)	(2.904)	(2.674)
CR cnrs	−0.362	−0.271	−0.235	−0.223	−0.015	0.136	−0.036	0.396
	(0.237)	(0.297)	(0.153)	(0.182)	(0.138)	(0.165)	(0.345)	(0.412)
CR inra	0.316^c	0.111	−0.201	−0.065	−0.053	-0.330^b	−0.192	-1.032^a
	(0.190)	(0.242)	(0.122)	(0.151)	(0.111)	(0.137)	(0.277)	(0.342)
DE ehess	0.198	0.515	−0.477	0.145	−0.596	-0.826^b	-2.066^b	-2.588^a
	(0.645)	(0.654)	(0.415)	(0.435)	(0.376)	(0.394)	(0.940)	(0.985)
DR cnrs	-1.074^a	-0.997^a	-0.448^a	-0.511^a	-0.395^a	-0.329^b	-1.395^a	-1.287^a
	(0.203)	(0.221)	(0.131)	(0.155)	(0.119)	(0.140)	(0.297)	(0.352)
DR inra	-0.969^a	-0.798^a	-0.730^a	-0.650^a	−0.138	0.134	-1.036^a	−0.169
	(0.177)	(0.231)	(0.114)	(0.160)	(0.103)	(0.145)	(0.258)	(0.363)
PR	-1.022^a	-0.916^a	-0.319^a	−0.132	-0.387^a	-0.364^a	-0.896^a	-1.194^a
	(0.154)	(0.197)	(0.099)	(0.140)	(0.090)	(0.127)	(0.225)	(0.320)
Insee	-0.924^b	-1.031^b	-0.985^a	-0.839^a	-0.817^a	-0.837^a	-2.434^a	-2.462^a
	(0.437)	(0.423)	(0.282)	(0.282)	(0.255)	(0.255)	(0.638)	(0.637)
IPC	0.127	0.070	-1.396^a	-1.140^a	-0.981^a	-1.021^a	-3.454^a	-3.301^a
	(0.336)	(0.343)	(0.217)	(0.323)	(0.196)	(0.293)	(0.490)	(0.740)
No-05	0.465^a	0.423^b	0.123	0.160	-0.218^a	−0.014	-0.522^a	−0.082
	(0.138)	(0.183)	(0.089)	(0.110)	(0.081)	(0.100)	(0.201)	(0.250)
No-37	0.261	1.520^c	0.168	0.578	1.607^a	0.879^c	2.324^b	1.806
	(0.800)	(0.844)	(0.516)	(0.510)	(0.467)	(0.462)	(1.167)	(1.149)
Time Fixed Effect	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.43	0.48	0.80	0.89	0.58	0.73	0.65	0.79
Observations	1209	1209	1209	1209	1209	1209	1209	1209

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. A-CR: other research fellow; A-MCF: non-university assistant professor; A-PR: non-university professor; CR cnrs: Cnrs research fellow; CR inra: Inra research fellow; DE ehess: Ehess research professor; DR cnrs: Cnrs research professor; DR inra: Inra research professor; PR: university professor; Insee: national statistical institute engineers; IPC: Ponts-et-Chaussées engineers. Reference: assistant professor.

Table 3.25: Variance analysis of the department fixed effects for the probability to publish

	Without individual F.E.		With individual F.E.	
	Stand. error	Correlation	Stand. error	Correlation
Explained: department fixed effect	0.660	1.000	0.584	1.000
Composition effects	0.254	-0.131	0.216	0.057
Size	0.085	0.128	0.008	0.011
Gender	0.025	-0.119	0.071	0.028
Age	0.117	-0.066	0.187	0.002
Positions	0.190	-0.176	0.178	0.055
Research strategy effects	0.667	0.814	0.383	0.479
Diversity	0.285	0.695	0.169	0.393
Research Access	0.026	-0.162	0.095	0.016
Non-USA Openness	0.084	0.368	0.002	-0.072
USA Openness	0.050	0.337	0.084	0.109
Heterogeneity	0.417	0.694	0.283	0.391
Stars	0.040	0.291	0.096	-0.043
Residuals	0.315	0.477	0.476	0.815

The table presents the variance analysis of the estimation reported in Table 3.23 column (1) and of the same regression when individual fixed effects (F.E.) are considered in the first step estimation, which is reported in Table 3.24 column (1) in 3.F. For the meaning of the figures reported, see the footnote in Table 3.5.

Table 3.26: Variance analysis of the department fixed effects for the publication quantity

	Without individual F.E.		With individual F.E.	
	Stand. error	Correlation	Stand. error	Correlation
Explained: department fixed effect	0.614	1.000	0.585	1.000
Composition effects	0.290	0.115	0.305	0.010
Size	0.169	0.272	0.137	0.254
Gender	0.053	-0.098	0.094	-0.023
Age	0.100	-0.020	0.123	-0.087
Positions	0.146	-0.038	0.191	-0.100
Research strategy effects	0.585	0.776	0.581	0.725
Diversity	0.193	0.647	0.228	0.653
Research Access	0.017	-0.213	0.117	-0.025
Non-USA Openness	0.095	0.415	0.097	0.338
USA Openness	0.012	0.358	0.049	0.304
Heterogeneity	0.393	0.673	0.338	0.579
Stars	0.075	0.333	0.112	0.283
Residuals	0.279	0.454	0.307	0.525

The table presents the variance analysis of the estimation reported in Table 3.23 column (3) and of the same regression when individual fixed effects (F.E.) are considered in the first step estimation, which is reported in Table 3.24 column (3) in 3.F. For the meaning of the figures reported, see the footnote in Table 3.5.

Table 3.27: Variance analysis of the department fixed effects for the publication top quality

	Without individual F.E.		With individual F.E.	
	Stand. error	Correlation	Stand. error	Correlation
Explained: department fixed effect	1.619	1.000	1.039	1.000
Composition effects	0.664	-0.166	0.445	-0.162
Size	0.147	0.056	0.049	0.002
Gender	0.127	-0.088	0.045	-0.063
Age	0.279	-0.026	0.198	-0.051
Positions	0.487	-0.205	0.389	-0.152
Research strategy effects	1.684	0.813	0.971	0.665
Diversity	0.445	0.572	0.328	0.540
Research Access	0.065	0.323	0.011	0.186
Non-USA Openness	0.322	0.508	0.124	0.332
USA Openness	0.204	0.490	0.182	0.347
Heterogeneity	0.923	0.721	0.540	0.600
Stars	0.372	0.445	0.141	0.276
Residuals	0.763	0.471	0.694	0.668

The table presents the variance analysis of the estimation reported in Table 3.23 column (7) and of the same regression when individual fixed effects (F.E.) are considered in the first step estimation, which is reported in Table 3.24 column (7) in 3.F. For the meaning of the figures reported, see the footnote in Table 3.5.

3.G. Individual first step regressions, T=t+2

Table 3.28: Determinants of individual publications, $\tau = t + 2$

	Publishing		Quantity		Quality		Top quality	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Women	-0.374^a	-0.375^a	-0.073^a	-0.195^a	-0.003	-0.220^a	-0.032	-0.953^a
	(0.03)	(0.02)	(0.01)	(0.06)	(0.02)	(0.08)	(0.05)	(0.21)
Age	-0.098^a	-0.085^a	-0.023^a	-0.059^a	-0.008	-0.058^a	-0.036^c	-0.244^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)
Age square	0.000^a	0.000^a	0.000^a	0.000^a	0.000	0.000^b	0.000	0.001^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Authors number			-0.944^a	-0.873^a	0.248^a	0.241^a	0.911^a	0.869^a
			(0.01)	(0.01)	(0.03)	(0.02)	(0.06)	(0.04)
Non-USA Openness			0.272^a	0.244^a	0.243^a	0.231^a	0.832^a	0.785^a
			(0.02)	(0.01)	(0.03)	(0.02)	(0.08)	(0.04)
USA Openness			0.267^a	0.258^a	0.364^a	0.310^a	1.097^a	0.939^a
			(0.03)	(0.02)	(0.03)	(0.02)	(0.10)	(0.06)
Star			0.397^a	0.334^a	1.274^a	1.196^a	3.641^a	3.374^a
			(0.03)	(0.02)	(0.04)	(0.03)	(0.10)	(0.07)
Diversity			-0.403^a	-0.297^a	-0.057^a	-0.063^a	-0.063	-0.078^a
			(0.01)	(0.01)	(0.02)	(0.01)	(0.04)	(0.03)
Specialisation				0.310^a		0.011		0.019
				(0.01)		(0.01)		(0.02)
Quantity					0.024^b	0.014	0.310^a	0.274^a
					(0.01)	(0.01)	(0.03)	(0.03)
Selection			-0.029	0.965^b	-0.319^a	1.243^b	-1.084^a	5.397^a
			(0.03)	(0.42)	(0.05)	(0.57)	(0.15)	(1.45)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Jel Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department Time FE	No	Yes	No	Yes	No	Yes	No	Yes
R^2			0.435	0.559	0.356	0.450	0.442	0.529
Observations	827221	783237	19022	19022	19022	19022	19022	19022

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. Authors number: average per article (in logarithms). Individual diversity: number of active Jel codes (in logarithms). Selection: inverse of Mills' ratio calculated with a Probit equation. Specialisation: share of department output in the Jel code (in logarithms). Quantity: number of papers discounted by the number of authors.

Table 3.29: Determinants of department fixed effects, $\tau = t + 2$

	Publishing		Quantity		Quality		Top quality	
	OLS	FGLS	OLS	FGLS	OLS	FGLS	OLS	FGLS
Size	-0.003	-0.002	0.139^a	0.122^a	0.033^c	0.031^c	0.015	0.035
	(0.014)	(0.014)	(0.012)	(0.012)	(0.018)	(0.018)	(0.043)	(0.043)
Women	0.293^a	0.286^a	0.043	-0.064	0.168	0.295^c	1.096^a	1.348^a
	(0.103)	(0.107)	(0.088)	(0.118)	(0.131)	(0.172)	(0.319)	(0.424)
Age	0.048^a	0.047^a	0.013^a	0.004	0.018^a	0.017	0.078^a	0.072^a
	(0.004)	(0.004)	(0.003)	(0.008)	(0.005)	(0.011)	(0.011)	(0.027)
Diversity	0.703^a	0.703^a	0.272^a	0.189^b	0.301^a	0.351^a	1.196^a	1.234^a
	(0.024)	(0.024)	(0.020)	(0.078)	(0.030)	(0.108)	(0.074)	(0.273)
Research Access	-0.016^c	-0.016^b	-0.006	0.000	0.011	0.003	0.018	-0.009
	(0.008)	(0.008)	(0.007)	(0.008)	(0.010)	(0.012)	(0.025)	(0.028)
Non-USA Openness	0.856^a	0.855^a	0.617^a	0.537^b	0.571^c	0.718^c	3.248^a	3.581^a
	(0.270)	(0.270)	(0.230)	(0.263)	(0.344)	(0.389)	(0.835)	(0.949)
USA Openness	0.789^b	0.784^b	-0.307	-0.391	1.759^a	1.743^a	4.710^a	4.877^a
	(0.369)	(0.369)	(0.316)	(0.314)	(0.472)	(0.475)	(1.143)	(1.145)
Heterogeneity	-1.318^a	-1.317^a	-0.704^a	-0.432^b	-0.569^a	-0.696^b	-2.543^a	-2.822^a
	(0.037)	(0.037)	(0.032)	(0.199)	(0.047)	(0.271)	(0.114)	(0.691)
Stars	1.043^a	1.043^a	0.914^a	0.666^b	1.198^a	1.441^a	5.832^a	5.762^a
	(0.333)	(0.333)	(0.284)	(0.334)	(0.425)	(0.493)	(1.030)	(1.205)
Positions	yes	yes	yes	yes	yes	yes	yes	yes
Time Fixed Effect	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.86	0.86	0.76	0.91	0.62	0.79	0.77	0.90
Observations	1152	1152	1152	1152	1152	1152	1152	1152

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

3.H. IV first steps regressions

Table 3.30: IV First step regressions

	Size (1)	Div. (2)	Div. (3)	Div. (4)	Het. (5)	Stars (6)
Population 1990	0.677[a]	0.188[a]		0.037		
	(0.06)	(0.02)		(0.03)		
% Engineers	6.274[a]		1.511[a]	1.592[a]	-1.241[a]	0.111[a]
	(0.87)		(0.39)	(0.40)	(0.21)	(0.01)
Peripherality		0.382[a]			-0.063	0.005[a]
		(0.05)			(0.04)	(0.00)
Population 1999			0.045			
			(0.03)			
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.416	0.081	0.046	0.045	0.034	0.128
Observations	735	735	735	735	735	735

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. Div. = diversity. Het. = heterogeneity.

Chapter 4

Gender Differences in a Micro Labour Market: Promotions amongst Academic Economists in France [1]

Differences in promotion across genders are still prevalent in many occupations. In this paper we examine promotion gaps using data for academic economists in France, where promotion occurs through a national contest. We test whether the gender gap is due to discrimination, to a higher cost of promotion for women, or to different preferences for salaries versus department prestige across genders. Our results find no evidence of any of these mechanisms, and raise the question of whether the promotion gap is due to women participating less in or exerting lower effort during contests.

1. This chapter has been co-written with Pierre-Philippe Combes and Cecilia Garcia-Peñalosa. It has benefited from the excellent research assistance of Philippe Donnay and Charles Laitong to whom we are very grateful.

4.1. Introduction

Despite the rapid increase in female educational attainment in the last decades of the 20th century, the labour market outcomes of men and women still differ substantially, whether in terms of wages or seniority. The literature on gender wage gaps is vast, but differences across genders in promotions have received much less attention.[2] This is partly due to the fact that promotion is based on a wide range of individual characteristics, many of which are unobservable to researchers, and hence it is difficult to assess the extent of differences across genders. In recent years, promotion gaps in academia have started to be examined in detail, both because there are clearly defined ranks that allow an unambiguous definition of what is a promotion, but also because the key consideration in actual promotion decisions, an individual's publications, can be observed and thus controlled for.

In this paper we use data for academic economists in France to measure the extent of promotion gaps across genders and to test alternative hypotheses of why they exist. Despite the growing number of empirical studies measuring gender gaps in promotion for academics, the reasons for these gaps are still poorly understood. The evidence indicates that part of the gap is due to women having a lower research output than men (see, for example, Ginther and Kahn (2004)), yet, a substantial fraction remains unexplained. We use the particular features of the French academic system to test competing hypotheses for the presence of promotion gaps, and in particular whether discrimination against women is present.

A key question concerning gender differences in labour markets is whether there is

2. For an overview of work on gender wage gaps see Blau and Kahn (2000). The literature on gender gaps in promotions stems from the seminal work by Lazear and Rosen (1990). For empirical evidence outside academia see, for example, Winter-Ebmer and Zweimuller (1997).

some form of discrimination against women or whether it is women that self-select themselves into certain jobs, occupations, or hierarchical levels because of, for example, family constraints. Using data for academics has an important advantage. Unlike in many private sector jobs, where a promotion is associated with longer hours and a requirement for greater availability outside normal working hours, academics have similar obligations and constraints at all hierarchical levels. Even if more senior academics tend to be involved in university administration and outside responsibilities such as participating in committees, seeking funding or performing editorial activities, these activities are to a large extent voluntary. They are not 'required' by the promotion and not performing them would not imply that the individual is demoted. Female assistant professors should thus not feel more constrained in terms of combining career and family duties by becoming full professors, and there is hence no obvious reason why they would prefer not to be promoted. Moreover, male and female academic economists are likely to have rather homogeneous labour market attachment, as argued by Kahn (1995), removing one of the reasons often branded to justify lower promotion rates for females.

In this context, we consider four potential causes of the unexplained component of promotions gaps: discrimination, higher costs of promotion for women, a different tradeoff between salaries and prestige across genders, and self-censorship in contests. The special features of the French academic system, such as a national salary scale, the need to go through a national contest in order to be promoted, and the existence of several categories of academics with different requirements upon promotion, allow us to tests the first three hypotheses.

The paper contributes to the literature using data on academics to try to understand gender gaps in labour markets, which dates back to the seminal work of Cole and Cole (1973). Early on, empirical analyses identified both lower wages and lower promotion rates for female economists; see Johnson and Stafford (1974) and Farber (1977). More recent work, such as Ginther and Hayes (1999), indicates that salary gaps are explained by differences in academic rank in the US. Yet, promotions to tenure and to full professor rank are still affected by gender even after controlling for research output and demographic characteristics. Some studies claim a decline in the promotion gap over time, while others find that it is large even in recent decades; see McDowell, Singell and Ziliak (2001) and Ginther and Kahn (2004). Evidence for the UK by Blackaby, Booth and Frank (2005) indicates that there are both gaps in promotions and in within-rank pay across genders, and their findings suggest that these are partly due to differences in the outside offers received by men and by women, while Sabatier (2010) documents the existence of a promotion gap in France.

Our contribution is twofold. On the one hand, we focus on an entirely different promotion system. Most of this literature has focused on the US and the UK, which have an academic labour market with much greater wage and promotion flexibility than those found in most European countries. Because we use data for academic economists in France, we can examine whether promotion gaps also exist in a labour market that operates in an entirely different way, with salaries being fixed at the country-wide level and promotions being decided by national committees and not by the department where the individual is employed. On the other hand, we can use some of the features of the French system to test different hypotheses in a way that would not be possible in the more flexible Anglo-Saxon academic systems. In France, there exist two types of academic positions -standard university professorships and pure research positions- which have different promotion rules. In both cases promotions are decided by a national committee; however, university positions require the individual to move to another university upon promotion while for pure research positions such mobility is not required.

Our results indicate that the promotion gap is prevalent and large in France, and rule out several potential causes of such gap. In particular we find no evidence that promotion differences across the sexes are due to discrimination, to women being affected by the mobility cost associated with promotion, or to different preferences concerning salaries and department prestige across the genders. This raises the question of whether the promotion gap is due to women exercising some form of self-censorship as far as promotions are concerned, either because they are less likely to seek promotion or because they exert less effort during the required contests. Such an explanation would be consistent with the patterns obtained in experiments in which women perform badly in competitions against men; see Gneezy, Niederle and Rustichini (2003).

The paper is organised as follows. Section 4.2 starts by examining the possible reasons why women are less likely to be promoted, we then describe the French academic system and explain how the different hypotheses can be tested. Section 4.3 describes the data, an exhaustive panel of academic economists in France in the period 1990 to 2008. Our results are presented in section 4.4, while section 4.5 concludes.

4.2. Why are there so few female professors?

4.2.1. Discrimination or self-selection?

Although a substantial literature has examined the promotion gap across genders, a clear explanation is still lacking. There are three reasons why women may be less likely to be promoted: discrimination, differences in costs or preferences across genders, and self-censorship in applying for promotions. Let us consider them in turns.

The first possible explanation is simply that those making the promotion decisions discriminate against women. This discrimination may be due to tastes or statistical, and implies that a promotion committee would, for given research output, prefer to promote a man.

Alternatively, women may make different choices from men that in turn result in different observed promotion probabilities. There may be costs associated with promotion, and if these are higher for women than for men, this could result in the former applying less often than men for a promotion. In France, promotions tend to require moving to a different department and thus tend to imply moving to a different city. If females' bargaining power in the household is lower, it may be harder for female than for male academics to impose the cost of moving on their families. Another reason why women are not promoted could be that they may have different preferences over department prestige and income. Suppose that women have a lower marginal utility of income because, often, they are the second earner in the household. They may then be willing to be in a more prestigious department even if it implies lower wages. Since in the US more prestigious departments also have tougher promotion thresholds (see McDowell et al. (2001)), this would result in a lower promotion probably for women. Lastly, there could be self-censorship. Women may choose not to apply for promotion simply because they are less inclined to compete in tournaments than men, as indicated by experimental evidence; see Gneezy et al. (2003).

Existing work has had difficulty in testing different hypotheses. Part of the problem is that the tradeoff that exists in the Anglo-Saxon academic system between prestige of the department and promotion thresholds makes it difficult to assess the alternatives. In contrast, the French system presents a number of features, such as a common national promotion system, that will allow us to evaluate the various options.

4.2.2. The French academic system

The French academic system has a number of features that we intend to exploit to test the above hypotheses. There are two types of academic positions in France. The most common are university positions, where the individual is a professor with a substantial teaching load. There exist also a number of public research instances, such as the Cnrs and the Inra, that have pure research positions.[3] Researchers in this category are hired by the Cnrs or Inra, who pay their salaries, but are attached to a university and are hence located in its economics department just like the university professors are. Researchers generally undertake some teaching, for which they get an extra salary, and participate in department life in the same way as standard professors.

For all types of position there is an entry level category, termed 'Rank B', which includes the 'maitre de conference' positions at the university and the 'chargé de recherche' at the other instances. The individual can then be promoted to 'Rank A', the equivalent to full

3. Centre National de la Recherche Scientifique, and Institut National de la Recherche Agronomique, respectively.

professor. Both rank A and rank B positions are tenured, which implies that an individual who does not get a promotion can spend her/his entire career in a 'junior', *i.e.* rank B, position. Because of this possibility, we will term the two types of positions 'rank A' and 'rank B', rather than junior and senior. The promotion from rank B to rank A entails a substantial salary increase, and especially a much steeper slope for salaries over time. The salary scales are identical for university professors and researchers and set by the Ministry of Higher Education and Research.[4]

Promotions take place through a national contest, a 'concours', and are thus not decided by the department in which the individual holds her/his current position.[5] Participation in this contest is public information, and at the end of it a list with the ranking of those that have undertook it is published. For university professors, departments publish before the start of the contest the positions that they have available. The candidates then choose, sequentially and starting with the highest ranked, which university to join. When promoted, the individual does not stay in the university where s/he held a rank B position and has to move to a different department. After three years s/he is allowed to move to another university, if the latter wishes to recruit her/him, including the university where s/he held the rank-B position. For researchers, academics that are promoted can choose to stay at the university where they are or move if another department wishes to recruit them. The university does not need to have an open position for them since the researchers' salary is always paid by the Cnrs, Inra or research institution to which s/he is affiliated.

4. Within each broad rank, A and B, there are subranks that affect salary. Subrank promotion is also decided by a national commitee, although the cost here is minimal, with application for a promotion requiring filling a form and submitting a vitae. Promotion to a higher subrank does not involve change of department.

5. Some internal promotions exists for individuals that have undertaken substantial administrative tasks at the university, but they are extremely rare. See Combes, Linnemer and Visser (2008) for more details.

It is obvious that promotion is very costly in the French system. There is a cost involved in all type of positions, since a concours needs to be passed which requires a substantial effort. The cost is much higher for university professors than for researchers, both because of the geographical mobility involved in being promoted to full professor, but also because the individual faces considerable uncertainty about where s/he will be eventually recruited if s/he passes the contest (since where s/he ends up depends on her/his ranking and hence on the entire pool of candidates). Since candidates seeking to become full professor are typically between 30 and 40 years of age, the process occurs at a moment in the life-cycle when family constraints are likely to be substantial. If women are less geographically mobile than men, then the cost will be greater for them.

Under the concours system, promotion is decided by a national committee. This aspect is extremely important for our purposes. In a system in which there is a positive correlation between prestige and promotion threshold women may choose to select into less prestigious department because promotion is easier in those, and hence the measured gap would underestimate the actual promotion gap. Members of the national committees are academic economists, drawn from various universities in the country and areas of expertise. Committees tend to be large and represent a wide spectrum of universities, not necessarily the most prestigious. Because members of these committees have to be of the full professor rank and because of the age distribution of the population of academics, there is a strong male dominance in these committees. Committees change regularly, every two to four years depending on the particular instance.

A second feature of the French system is that there is no trade-off between department prestige and salaries.[6] Academic salaries are determined according to a national scale based on rank, and rank is, as we saw, decided at the national level.[7] In the French system, there is thus no reason to prefer being employed in a less prestigious department. Salaries are the same across universities and the threshold for promotion is set nationally. Hence, if they could, all men and women should prefer being in a more prestigious department.

4.2.3. Testing the different hypotheses

We can use the above features of the French academic system to confront the various hypotheses about lower promotion rates for women. In order to do so, we will examine both the determinants of promotion and of the probability of being in a prestigious department. We consider two categories of department, somewhat equivalent to the division in the US between the top-50 and other departments, and define prestigious departments as those that have the largest research output in international journals (see Appendix 4.A for details). For this reason, we will term the most prestigious departments 'international' departments and the rest 'national' departments. Our tests will proceed as follows:

Discrimination: We undertake two tests for the presence of discrimination. The first consists of looking at the effect of gender on the probability of being promoted and the probability of being in an international department. As argued by Lazear and Rosen (1990) it is hard to understand why women would be discriminated against in promotions but not in other labour market experiences. In fact, in many instances the evidence tends to indicate that women have a lower probability of both being promoted and being hired,

6. Early work on the US indicated that there was a tradeoff between salary and prestige; see Hansen et al. (1978).

7. A few of the international departments pay, out of their own funds, an extra salary on top of the one paid by the university/cnrs/etc. This practice is, however, restricted to only a few members in a handful of departments.

as for example in the case of top US orchestras; see Goldin and Rouse (2000).

For our purposes, we exploit the fact that those making hiring and promotion decisions are always academic economists. In the case of pure research positions, there is a national committee that makes both the hiring and the promotion decision. If these individuals discriminated against women, we would expect them to do so both in promotions and hiring, and hence we should find that women are both less likely to be promoted and less likely to be in an international department. For university positions, junior hiring is decided locally (by the department) while promotions are decided by a national committee. At any point in time the individuals making these two decisions are hence not the same. However, the national committee is chosen from current senior academics economists, that is, it is drawn from the pool of those making the junior recruitment decisions. It is in principle possible that there is a selection process such that even though on average academic economists do not discriminate against women, those that participate in the national committee do discriminate. Given the high turnover of the members of the national committee we find this unlikely and hence expect to find the same attitudes towards women both at the department and at the national committee level.

The literature on gender differences in labour markets has extensively focused on distinguishing between the effect of differences across the sexes in characteristics and in the coefficients associated with those characteristics, with the latter being often interpreted as an indication of discrimination; see Oaxaca (1973). Our second test for discrimination is hence to look at the difference in returns to publications between men and women. If there were discrimination against women, we would expect to find smaller returns to publications for the latter.

Higher cost of promotion: In the US system, it is not clear why women may have different costs of promotion. The nature of academia and the substantial flexibility associated with this type of jobs implies that jobs at the junior or senior level are equally likely to be compatible with family life. As a result, there is no reason why women would prefer not to be promoted. In contrast, in the French system the mobility required (of professors but not researchers) if promoted implies that there is a very specific cost: the need to move to another university and hence, in most cases, to another city. It is likely that this cost is, on average, higher for women than for men, implying that the latter are less likely to apply for promotion. To test this hypothesis we can exploit the difference between the implications of promotion for standard professors and for researchers. If women were not trying to get promoted because of the cost of mobility, then we expect to find a negative and significant effect of being female on the probability of promotion for professors but no effect for researchers.

Differences in preferences about prestige and salary: We can also exploit the difference between promotion implications for professors and researchers to test the hypothesis of different preferences over salary and prestige across genders. Since for university professors being promoted to a senior position requires changing department, promotion brings a higher salary but also the risk of having to move to a less good department. If women are less willing to trade-off prestige for income, then those who hold a junior position in an international department may not want to apply for promotion. In contrast, for junior researchers promotion does not require a change of department, and there is no reason for women not to seek promotion even if they are in an international department.

We can then examine whether there are differences in the effect of gender on promotion for women depending on the type of job that they hold. If women have a stronger preference for department prestige and/or a lower one for income, they will have a higher probability of being rank B in an international department and a lower probability of being rank A in national departments than men. In contrast, because promotion does not require changing department for researchers and we would expect to find no differences in the probability of being rank A or rank B across genders. In order to do so, we can use a multinomial model with four outcomes: being rank B in a national department, rank B in an international department, rank A in a national department and rank A in an international one.

Self-censorship: There is nothing in the data that allows us to test for self-censorhip, whether in terms of not applying for promotion or exerting low effort during the concours. We hence view self-censorship as our default explanation for the gender gap in promotions.

To sum up, we intend to perform the following tests:

- Discrimination I: We compare the sign of the gap women-men in promotion and the gap women-men in being in an international department. Discrimination against women on the part of promotion and hiring committees would imply that both gaps are negative.
- Discrimination II: We consider the difference in returns to publications between women and men. A lower return to publications for women would be evidence of discrimination against women on the part of promotion and hiring committees.
- Higher cost of promotion: We consider the difference between the gender gap in promotions for university professors and for researchers. If the mobility associated with promotion were the reason why women seek promotion, there would be an impact of gender for professors but not for researchers.

– Different preferences about prestige and salary: In a multinomial model with four outcomes, we compare the coefficients on gender for professors and researchers. A difference in preferences would result in gender having a positive (negative) effect of being rank B in an international department (rank A in a national department) for professors. Gender should have no effect on these probabilities for researchers.

In all cases we will consider the equations in levels rather than running differences-in-differences because, as well as testing for the above hypotheses, we are interested in understanding the overall process determining promotion and hiring for academic economists in France, and hence on the magnitude and significance of the effects of variables other than gender.

4.2.4. Empirical specification

We suppose that the probability of being in each category is given by

$$\Pr(S,i) = \Phi\left(X_i'\beta\right) \tag{4.1}$$

where the two states S are being rank A or not, and $\Phi(.)$ denotes the logistic function. The term $X_i'\beta$ in equation (4.1) is

$$\begin{aligned} X'\beta_i = \alpha + \alpha^f\ \delta_i + \beta_1\ Age_i + \beta_2\ Age_i^2 + \beta_3\ Pub_i \\ + \beta_4\ Pub_i \times Quantity_i + \beta_5\ Pub_i \times Quality_i \end{aligned} \tag{4.2}$$

implying that the probability of promotion is a function of age Age_i, whether or not the individual has published in Econlit-classed journals (*i.e.* whether s/he is a 'publisher' measured by the dummy Pub_i), the number of publications and the average quality of these publications, denoted respectively $Quantity_i$ and $Quality_i$, both measured in logs (see below for the exact measurement). The variable δ_i is a dummy taking the value 0 for males and 1 for females, so that α^f measures the differences in promotion probability for men and women with the same characteristics. We also control for individual's fields of specialisation (defined by JEL codes) and whether s/he belongs to Cnrs, Inra or other research instance rather than being a standard university professor. All regressions include time fixed effects.

Alternatively, we may suppose that some of the coefficients differ across genders and estimate the model with

$$\begin{aligned} X'\beta_i = {} & \alpha + \alpha^f\ \delta_i + \beta_1\ Age_i + \beta_2\ Age_i^2 + \beta_3\ Pub_i \\ & + \beta_4\ Pub_i \times Quantity_i + \beta_5\ Pub_i \times Quality_i \\ & + \beta_3^f\ Pub_i + \beta_4^f\ Pub_i \times Quantity_i + \beta_5^f\ Pub_i \times Quality_i \end{aligned} \tag{4.3}$$

This expression allows for different returns to research output across genders, as it is possible that women have a higher or a lower return to publications.

The same model is run for the probability of being in an international department. Obviously, the qualities that lead to promotion in academia are also those that make an individual be recruited by an international department. Lastly, we also consider a multinomial model in which there are four outcomes, depending on whether the individual holds a junior or senior position, and whether s/he is in a national or international department. Note that it is not possible to look at the probability of promotion conditional on being in, say, an international department, since in France promotion and affiliation tend to be simultaneously determined.

4.3. The Data

4.3.1. Measuring research output

We measure the research output of individual i at date t by its cumulative publication record at date t. Publication records are measured as weighted sums of publications. All publications come from the EconLit database, which includes more than 560,000 papers published in more than 1200 journals between 1969 and 2008. Three dimensions enter the weighted scheme of publications: the quality of journals, the number of authors and the relative number of pages.

We measure the quality of publications using the journals weighting scheme proposed by Combes and Linnemer (2010). Two different degrees of convexity in the distribution of journals' weights have been proposed by Combes and Linnemer, and we use the most convex one (*i.e.* the one that most values quality), but our results are unchanged when we use the least convex one. We also consider the number of authors of each publication, and assume constant returns to scale at the coauthor level, so that we divide each publication by the number of authors.[8] Lastly, since many journals now publish notes and short papers, we weight by the number of pages to capture the idea that longer articles contain

8. This assumption is useful for aggregation purposes, since when measuring the output of a department as the sum of outputs of its members, we would not want a system such that an article written by two members of the department accounts for more or less than the same article written by only one member. Assuming increasing returns to scale within the coauthors team would lead to underestimate the output of a co-written article whereas assuming decreasing returns to scale would overestimate it. See Bosquet and Combes (2012) for further discussion.

more ideas. Nevertheless, since the layout can be different from one journal to another, we consider an article's length relative to the average length in that journal. Then, the ratio of the number of pages of article a over the average number of pages of articles published in the journal is used as weight. Because editors' policies can vary over time, the mean number of pages is calculated for each year separately.

The output of a researcher i at date t is then a weighted sum of her/his articles a published between 1969 and date t:

$$y_{it} = \sum_{a \in [1969,t]} \frac{W(a)}{n(a)} \frac{p(a)}{\bar{p}} \tag{4.4}$$

where $p(a)$ is the number of pages of article a, $\bar{p}$ is the annual average number of pages of articles in the journal, $n(a)$ the number of authors of the article, and $W(a)$ the weighting scheme for journals. Then, each individual receives three scores: a dummy equal to 1 if s/he has at least on publication in an EconLit-listed journal, the quantity of single-author-equivalent published articles, and the average equality of her/his articles, defined as y_{it} divided by the number of articles.

In our regressions we also include the individual's fields of specialisation. This is potentially an important control because men and women may choose to work in different fields; see Dolado, Felgueroso and Almunia (2008). If the former tend to choose more fashionable or prestigious areas of research than women, this may make promotion easier for men, and previous work has indicated the importance of controlling for field; see McDowell et al. (2001). We define 18 fields following the JEL broad field classification, and use the JEL code information on each published paper to create for each individual 18 variables that measure the share of her/his papers published in each JEL category.

4.3.2. Population and descriptive statistics

In order to get the entire population of French academic economists, EconLit is first merged with the list of academics provided by the French Ministry of Higher Education and Research, the Cnrs and Inra for the years 1990 to 2008. We merge publications by surname and initial, and then correct manually for those individuals having the same name and initial. For each individual we have information on age, rank, publication stock and department. We keep only individuals that are in departments larger than 4 full-time equivalent academics, which removes economists that are isolated in universities without real economics departments. Those (few) individuals for whom some of the individual characteristics (age or position for instance) are missing are also excluded.

Table 4.1: Descriptive statistics of the appartenance of researchers, panel 1990-2008

Belonging	Individuals	Observations	%	% Women	% A-rank
University	2510	30312	81.5	25.5	34.0
Cnrs	343	3608	9.7	22.0	41.4
Inra	238	2621	7.0	22.0	51.4
Other	91	674	1.8	13.6	100.0
Total	3182	37215	100.0	24.7	37.2

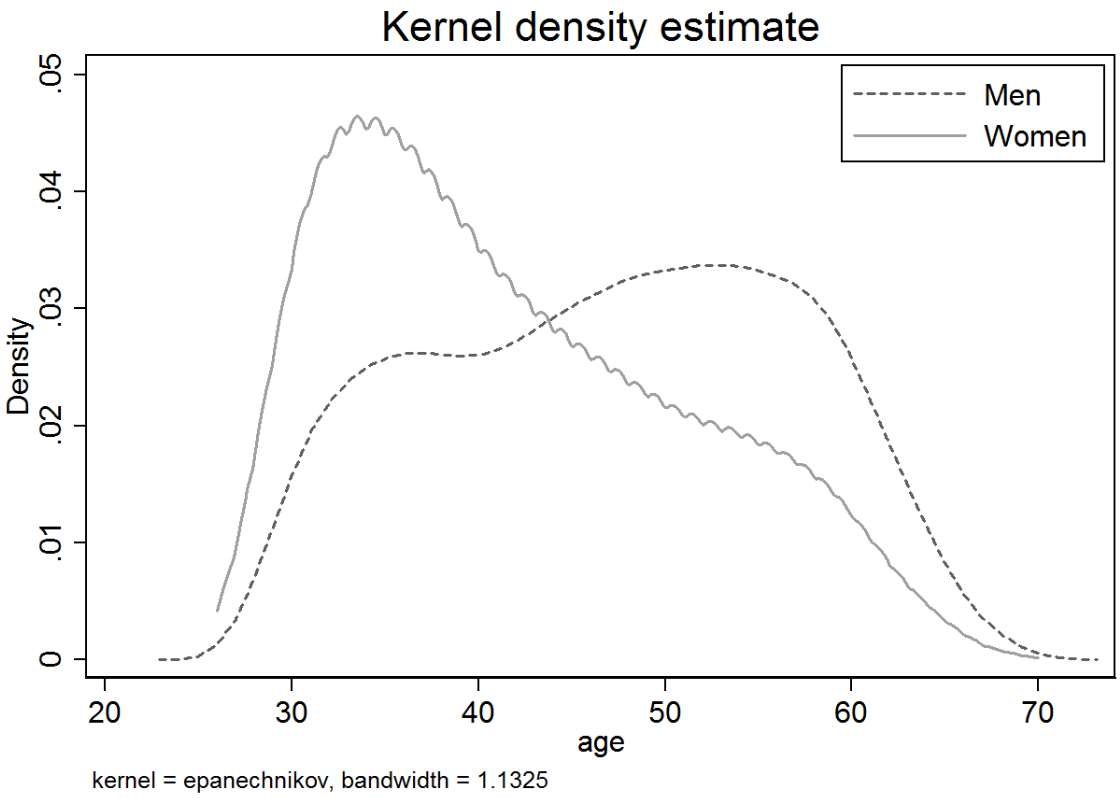
Cnrs = Centre national de la recherche scientifique. Inra = Institut national de la recherche agronomique. Other includes 'Institut national de la statistique et des études économiques' and 'Ponts-et-Chaussées' engineers as well as research directors of 'Ecole des Hautes Etudes en Sciences Sociales'.

Table 4.1 gives the decomposition of our sample in terms of institutional affiliation. We have 37,215 observations, spanning over 19 years and including a total of 3,182 academics. The panel is unbalanced as individuals enter and exit academia over our sample period, with the average number of observations per academic being 12 years. As we can see, the vast majority of observations (over 30,000) are for university professors, with researchers accounting for 81.5% of the total. Women account for 25% of observations, and they are overrepresented amongst university professors and underrepresented amongst researchers. This is not surprising given that obtaining a position as a researcher tends to require a stronger publication record and, as we will see below, women tend to have a weaker one than men. Slightly over a third of the population hold a rank A position, with the fraction being smaller for university professors (34%) and higher for researchers. Note that the data do not seem to indicate that selection is a problem that we should be concerned with. If women chose a career path that offered higher average promotion rates, ignoring selection would underestimate the impact of gender on promotions. Our data indicates that feminisation is lower for researcher, which also have a higher promotion rate, indicating that there is no selection of this type taking place.

Table 4.2 reports some descriptive statistics. The fraction of male academics that are rank A is 43%, while the figure is much lower for women: only 18%. The table indicates that some of this difference is likely to be explained by differences in observable characteristics of men and women. Women are on average 5 years younger than men in the sample. This is not surprising since the fraction of women in the profession has been increasing steadily over time, and this implies rather different age structures across genders, as can be seen from the smoothed density for the two groups depicted in Figure 4.1. As a result, the female population is younger and hence has a lower probability of having been promoted. A second important difference lies in publication records. The probability of publishing in EconLit journals is 67% for men but only 59% for women, while the average quantity and quality of publications is twice as high for men as for women.[9]

9. Obviously some of the differences in research output are also due to the age structure of the two populations, but Bosquet and Combes (2012) show that even when controlling for age women have lower

Figure 4.1: Smoothed densities of age distributions for men and women



Tables 4.3 and 4.4 report the position of women, both in terms of promotions and of affiliation to top departments. As we have seen, the promotion gap is large (25 percentage points), and Table 4.3 indicates that it is substantially smaller for university professors, for whom it amounts to 23 points, than for researchers, for whom it is 30 points. In contrast, when we look at international departments no such gap appear. The overall fraction of academics on top departments is 33%, with the figure being slightly higher for women than for men (35% versus 33%). The same pattern appears for the professor and researcher categories, with women having systematically a slightly higher probability of being in an

publication's productivity.

Table 4.2: Descriptive statistics by gender

	Observations	Minimum	Maximum	Mean	Standard error
	Total sample				
Age	37215	24	85	45.93	9.98
Prob. rank-A	37215	0	1	0.37	0.48
Women	37215	0	1	0.25	0.43
Prob. Int. Detp.	37215	0	1	0.33	0.47
Quantity	37215	0	101.7	2.59	4.95
Quality	37215	0	152.2	1.18	5.45
Publisher	37215	0	1	0.65	0.48
	Women				
Age	9200	26	70	41.93	9.50
Prob. rank-A	9200	0	1	0.18	0.39
Prob. Int. Detp.	9200	0	1	0.35	0.48
Quantity	9200	0	24.0	1.36	2.12
Quality	9200	0	62.5	0.70	3.84
Publisher	9200	0	1	0.59	0.49
	Men				
Age	28015	24	85	47.24	9.78
Prob. rank-A	28015	0	1	0.43	0.50
Prob. Int. Detp.	28015	0	1	0.33	0.47
Quantity	28015	0	101.7	2.99	5.51
Quality	28015	0	152.2	1.34	5.88
Publisher	28015	0	1	0.67	0.47

Int. Dept. = International Department. Productivity measures (quantity and quality) are in levels. We take their logs in the regression analysis.

Table 4.3: % A-rank by gender

A-rank	Total	%	Women	%	Men	%
	Total sample					
0	23386	62.8	7500	81.5	15886	56.7
1	13829	37.2	1700	18.5	12129	43.3
	University professors					
0	19998	66.0	6432	83.1	13566	60.1
1	10314	34.0	1305	16.9	9009	39.9
	Researchers					
0	3388	49.1	1068	73.0	2320	42.6
1	3515	50.9	395	27.0	3120	57.4

international department.

Table 4.4: % International Department by gender

Int Dept	Total	%	Women	%	Men	%
	Total sample					
0	24811	66.7	5994	65.2	18817	67.2
1	12404	33.3	3206	34.8	9198	32.8
	University professors					
0	21142	69.7	5234	67.6	15908	70.5
1	9170	30.3	2503	32.4	6667	29.5
	Researchers					
0	3669	53.2	760	51.9	2909	53.5
1	3234	46.8	703	48.1	2531	46.5

4.4. Results

4.4.1. Testing for discrimination

We start by examining the determinants of being rank A for the entire population, which are reported in Table 4.5. In column (1), only time fixed effects and gender are included in the logit model, with the dummy being equal to 1 for women. The marginal effect on gender is significant at the 1% level and large, implying an odds ratio of 0.30. [10] Including age reduces the effect on gender from -0.246 to -0.174, indicating that a large fraction of the difference in promotion is indeed due to the fact that the sample of women is younger than that of men. Column (3) includes our three measures of research output, whether or not the individual publishes, the quantity of publications and their quality. All three are highly significant and increase the probability of promotion.

Once we control for research output, the effect of gender falls to about a third of the initial one, indicating that to a large extent the lower promotion rate for women is due to

10. $exp(-1.205)$, -1.205 being the coefficient associated to the -0.246 marginal effect.

Table 4.5: Probability to hold a A-rank position, panel 1990-2008, marginal effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	logit	logit	logit	logit	logit	logit	probit	OLS
Women	-0.246[a]	-0.174[a]	-0.072[a]	-0.070[a]	-0.133[a]	-0.124[a]	-0.110[a]	-0.041[a]
	(0.004)	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)	(0.008)	(0.007)
Age		0.020[a]	0.015[a]	0.016[a]	0.017[a]	0.017[a]	0.016[a]	0.015[a]
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2		-0.000[a]	0.000	0.000	0.000	0.000	0.000	0.000[a]
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Publisher (Pub)			0.302[a]	0.259[a]	0.245[a]	0.260[a]	0.258[a]	0.271[a]
			(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
Pub*Quantity			0.143[a]	0.143[a]	0.141[a]	0.144[a]	0.148[a]	0.177[a]
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pub*Quality			0.036[a]	0.031[a]	0.031[a]	0.032[a]	0.032[a]	0.026[a]
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Women*Pub					0.073[a]	0.060[a]	0.050[a]	-0.005
					(0.017)	(0.017)	(0.015)	(0.014)
Women*Pub*Quantity					0.018[b]	0.020[b]	0.011	-0.006
					(0.008)	(0.008)	(0.008)	(0.007)
Women*Pub*Quality					-0.002	-0.002	-0.002	0.007[b]
					(0.003)	(0.003)	(0.003)	(0.003)
Specialisation	No	No	No	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	No	No	No	No	Yes	Yes	Yes
pseudo-R^2	0.041	0.118	0.336	0.350	0.351	0.362	0.360	
Observations	37215	37215	37215	37215	37215	37215	37215	36541
log-likelihood	-23555	-21648	-16296	-15965	-15933	-15664	-15713	-15920

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively.

them having published less. The effect is nevertheless still strong: being a woman reduces the probability of promotion almost as much as having one rather than two single-authored publications. Column (4) indicates that including the field of specialisation does not have much impact, neither on the effect on gender nor on those on age and output.[11] The effects we obtain are comparable to those found for the US. Ginther and Kahn (2004) find a raw gender gap of -0.213, which falls to -0.130 once age and publication records are included.

Column (5) allows for a different impact of research output for men and women. The interacted terms are positive and significant both for whether individuals publish and for

11. See Dolado et al. (2008) on gender gaps in specialisation amongst academic economists.

the quantity of publications, though not for their quality. In contrast the effect on gender falls, indicating that the now much lower probability for women to be promoted is partly offset by a higher return to research output. The overall impact of being a woman, evaluated at sample means for research output, is -0.067. We next include dummies for whether the individual holds a post other than a university position (*i.e.* is a researcher employed by the Cnrs, Inra, etc.), as different institutions may have different overall promotion rates. Although the coefficients on the dummies are significant, indicating such differences (the corresponding marginal effects are not reported, and we term this set of dummies 'belonging'), there is barely no impact on the measured effects of other variables. Columns (7) and (8) report the estimations of a probit and an OLS model, respectively, and indicate that the result that women are less likely to be promoted although they experience somewhat stronger returns to output are robust to the use of those specifications.

Table 4.6 presents the logit results for whether or not the individual is in an international department, using the same explanatory variables as before. As we saw earlier, women are more likely to be in a prestigious department, and the odds ratio either for the regression with only gender and for the one in which age is included give an odds ratio of 1.09, indicating that women are 2% more likely to be in an international department.[12] The odds ratio increases to 1.48 when output is included (column 3), reflecting the fact that women tend to be less productive, and falls to 1.31 when we interact research output with the gender dummy (column 5).[13] There seems to be no evidence of a significant difference in the effect of the quantity of publications across genders, although the effect of quality

12. $exp(0.086)$, 0.086 being the coefficient associated to the 0.019 marginal effect.
13. $exp(0.393)$ and $exp(0.269)$, 0.393 and 0.269 being the coefficients associated to the 0.079 and 0.053 marginal effects, respectively.

Table 4.6: Probability to be in an international department, panel 1990-2008, marginal effects

	(1) logit	(2) logit	(3) logit	(4) logit	(5) logit	(6) logit	(7) probit	(8) OLS
Women	0.019^a	0.019^a	0.079^a	0.072^a	0.053^a	0.048^a	0.047^a	0.043^a
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)
Age		-0.008^a	-0.008^a	-0.007^a	-0.007^a	-0.008^a	-0.008^a	-0.007^a
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2		0.000^a	0.000^a	0.000^a	0.000^a	0.000^a	0.000^a	0.000^a
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Publisher (Pub)			0.225^a	0.214^a	0.211^a	0.186^a	0.191^a	0.208^a
			(0.008)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)
Pub*Quantity			0.061^a	0.061^a	0.062^a	0.056^a	0.055^a	0.055^a
			(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Pub*Quality			0.055^a	0.051^a	0.054^a	0.048^a	0.050^a	0.052^a
			(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Women*Pub					0.001	0.018	0.017	0.024
					(0.016)	(0.016)	(0.016)	(0.016)
Women*Pub*Quantity					-0.007	-0.009	-0.007	-0.004
					(0.008)	(0.008)	(0.008)	(0.008)
Women*Pub*Quality					-0.012^a	-0.009^a	-0.009^a	-0.008^b
					(0.003)	(0.003)	(0.003)	(0.003)
Specialisation	No	No	No	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	No	No	No	No	Yes	Yes	Yes
pseudo-R^2	0.001	0.002	0.087	0.100	0.101	0.126	0.126	
Observations	37215	37215	37215	37215	37215	37215	37215	37215
log-likelihood	-23673	-23639	-21631	-21315	-21303	-20698	-20692	-21722

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

interacted with gender is negative and significant. As before, the results are replicated when we estimate a probit or an OLS model. The effect of gender on affiliation is substantial. In our preferred specification, column (6), the effect of being a woman on the probability of being in an international department is large. It is equivalent to increase the number of single-authored publications from 1 to the sample average, 2.6, or to increase the quality of publications from its average of 1.9 to 5.2, *i.e.* by two thirds of its standard deviation.

The positive effect of being female on the probability of being in an international department may have three causes. One possibility is that there is positive discrimination, so that departments consciously try to increase their female faculty and more prestigious ones manage to attract more women. Another is that there are unobservable abilities, *i.e.* abilities other than those measured by research output such as organisational skills or teaching ability, and that the population of female academics has, on average, higher unobservable skills than that of men. Higher average skills could be due to women spending more time in teaching and administration, which in turn could explain their lower research output. Also, because there are fewer women than men that enter academia, they have been drawn from the upper tail of the ability distribution and this may be reflected in higher non-publication skills; see Petrongolo and Olivetti (2008) on selection and gender. In this case, international departments would simply be rewarding a productive trait that is unobservable for us but not for them. This mechanism is to some extent supported by the evidence provided in Bosquet and Combes (2012) that women exert a positive externality on their colleagues, with individuals in departments with a higher proportion of women tending to have a higher research output. Lastly, the feminisation of international departments could be the results of the increasing practice of joint offers. Academic economists often have spouses in the profession, and hence negotiate joint appointments. International departments competing for the best academics may try to attract the husband by offering a job to the wife too, thus raising the probability of women being in a top department. [14]

When we compare the effects of quantity and quality across the two dependent variables we find that although being a publisher has similar effects on the probabilities of being promoted and on that of being in an international department (0.259 and 0.214 in the fourth columns of Tables 4.5 and 4.6), quantity matters more for promotion than for affiliation

14. In principle it would be possible to test for the latter hypothesis by using information on couples. Such information does not at the moment exist, although it is conceivable to run a survey in the future in order to obtain it.

(0.143 versus 0.061) while quality matters more for affiliation than for promotion (0.051 against 0.031). This is obviously consistent with the more selective nature of the more prestigious departments.

Tables 4.5 and 4.6 tell a surprising story: being female decreases an individual's probability of being promoted, yet it increases her/his probability of being in a prestigious department. The magnitude of the two effects is roughly the same, with the odds ratios being 0.60 and 1.44, respectively, for the two specifications in column (4) of the tables.[15] The evidence on hiring makes it difficult to see the lower promotion rate for women as being due to discrimination given that selection committees tend to be composed by the same individuals in the two cases, and there is no reason why they would choose to discriminate in one dimension but not in the other. The absence of discrimination is also supported by the fact that women experience higher returns to publications in the probability of being of rank A.

4.4.2. Alternative hypotheses

4.4.2.1. Mobility as a cost of promotion

As discussed above, a particular feature of the French academic system is that being promoted to rank A requires a change of university and hence, in most cases, requires moving to a different location. In contrast, for researchers promotions do not have such mobility requirements. We examine whether the required mobility is the reason why we observe lower promotion for women by comparing the coefficients on gender for university

15. $exp(-0.511)$ and $exp(0.362)$, -0.511 and 0.362 being the coefficients associated to the -0.076 and 0.072 marginal effects, respectively.

professors and for researchers.

Table 4.7: Individual regressions, panel 1990-2008, marginal effects

	Logit				Probit	
	University		Researchers		University	
	(1) A-rank	(2) int.dept	(3) A-rank	(4) int.dept	(5) A-rank	(6) int.dept
Women	-0.055[a]	0.074[a]	-0.084[a]	0.076[a]	-0.056[a]	0.072[a]
	(0.005)	(0.007)	(0.012)	(0.012)	(0.005)	(0.007)
Age	0.013[a]	-0.009[a]	0.039[a]	0.000	0.013[a]	-0.009[a]
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Age^2	0.000[a]	0.000[a]	-0.000[a]	-0.000	0.000[a]	0.000[a]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Publisher (Pub)	0.313[a]	0.209[a]	0.133[a]	0.083[a]	0.311[a]	0.214[a]
	(0.010)	(0.011)	(0.016)	(0.016)	(0.009)	(0.010)
Pub*Quantity	0.153[a]	0.052[a]	0.128[a]	0.048[a]	0.154[a]	0.052[a]
	(0.003)	(0.004)	(0.006)	(0.006)	(0.003)	(0.004)
Pub*Quality	0.039[a]	0.046[a]	0.020[a]	0.038[a]	0.038[a]	0.048[a]
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Specialisation	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	No	Yes	Yes	No	No
pseudo-R^2	0.383	0.065	0.409	0.362	0.383	0.065
Observations	30312	30312	6903	6903	30312	30312
log-likelihood	-11993	-17373	-2826	-3046	-11994	-17375

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively.

The results are reported in Table 4.7. Column (1) and (3) report the logit of being rank A for university professors and researchers, respectively. In order to make the coefficients on gender easier to interpret, in these and all subsequent regressions we do not include the interaction between gender and publications. The effects on gender are -0.055 and -0.084, but we cannot reject the null hypothesis of equality of these effects across the two groups. That is, the fact that promotion is associated with mobility for the majority of academics does not seem to be the reason why females experience lower promotion rates. Columns (2) and (4) report the logit results for affiliation to an international department for the two groups. As before, we find that women are more likely to be in prestigious departments and we cannot reject the null that the effects are equal across professors and researchers. These results are confirmed by the difference-in-differences regression reported in Table 4.15 in Appendix 4.C.

Table 4.8: Individual regressions, panel 1990-2008, marginal effects

	Total sample		University		Researchers	
	(1)	(2)	(3)	(4)	(5)	(6)
	A-rank	int.dept	A-rank	int.dept	A-rank	int.dept
Women	0.064^a	0.048^b	0.060^b	0.046^c	0.156^a	0.083^c
	(0.023)	(0.021)	(0.025)	(0.024)	(0.048)	(0.042)
Age	0.020^a	-0.009^a	0.015^a	-0.010^a	0.044^a	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)
Age^2	-0.000	0.000^a	0.000^b	0.000^a	-0.001^a	-0.000^b
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age*Women	-0.009^a	-0.001	-0.006^b	0.000	-0.024^a	-0.010^b
	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.005)
Age^2*Women	0.000	0.000^b	0.000	0.000	0.000^a	0.000^a
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Publisher (Pub)	0.271^a	0.193^a	0.312^a	0.209^a	0.134^a	0.082^a
	(0.008)	(0.009)	(0.010)	(0.011)	(0.016)	(0.016)
Pub*Quantity	0.148^a	0.054^a	0.154^a	0.052^a	0.128^a	0.050^a
	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)	(0.006)
Pub*Quality	0.032^a	0.046^a	0.039^a	0.046^a	0.021^a	0.038^a
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Specialisation	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	Yes	Yes	No	No	Yes	Yes
pseudo-R^2	0.363	0.127	0.386	0.065	0.412	0.366
Observations	37215	37215	30312	30312	6903	6903
log-likelihood	-15643	-20690	-11943	-17364	-2815	-3024

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

A possible reason why men and women have different promotion probabilities is that, as is often found in the literature on gender wage gaps, they have different returns to experience. In order to see whether this effect is important we interact age with gender, and run the logit regressions for promotion for all academics as well as for university professors and researchers. The results are reported in Table 4.8. The overall impact of gender is negative in the three cases for all except very young women, and stronger for researchers than for university professors, indicating again that the cost of mobility is not the reason why women experience a lower promotion rate.[16]

4.4.2.2. Multinomial regressions

We turn next to multinomial regressions of the status of individuals. We have two possible labour market outcomes in academia: the rank, which can be A or B, and the prestige of the department, national or international. We hence divide our population into four groups according to their status: national rank B, international rank B, national rank A, and international rank A. We then estimate a multinomial logit model for these groups.

As we discussed earlier, if women have a stronger relative preference for department prestige, they may choose not to apply for promotion in order to stay in an international department. Since promotion implies mobility for university professors, we expect women to have a higher probability of being rank B in an international department than men, and a lower probability of being rank A in national departments than men. In contrast, for researchers promotion does not require changing department and we expect to find no differences in the probability of being rank A or rank B across genders.

Table 4.9: Multinomial regressions, panel 1990-2008, marginal effects

	Total sample		University		Researchers	
	(1)	(2)	(3)	(4)	(5)	(6)
B-rank in a national department						
Women	0.145^a	0.006	0.135^a	-0.002	0.157^a	0.020^c
	(0.006)	(0.005)	(0.007)	(0.006)	(0.015)	(0.012)
Age		-0.006^a		-0.003^b		-0.018^a
		(0.001)		(0.001)		(0.002)
Age^2		-0.000^a		-0.000^a		0.000^a
		(0.000)		(0.000)		(0.000)
Publisher (Pub)		-0.315^a		-0.351^a		-0.163^a
		(0.008)		(0.010)		(0.015)

continued on next page

16. Recall that age is measured as true age minus 25.

continued from previous page

	Total sample		University		Researchers	
	(1)	(2)	(3)	(4)	(5)	(6)
Pub*Quantity		-0.118^a		-0.120^a		-0.099^a
		(0.004)		(0.004)		(0.007)
Pub*Quality		-0.048^a		-0.050^a		-0.036^a
		(0.002)		(0.002)		(0.003)
B-rank in an international department						
Women	0.102^a	0.058^a	0.094^a	0.057^a	0.147^a	0.064^a
	(0.006)	(0.005)	(0.006)	(0.006)	(0.014)	(0.011)
Age		-0.012^a		-0.010^a		-0.020^a
		(0.001)		(0.001)		(0.002)
Age^2		0.000^a		0.000^a		0.000^a
		(0.000)		(0.000)		(0.000)
Publisher (Pub)		0.056^a		0.045^a		0.036^b
		(0.007)		(0.008)		(0.014)
Pub*Quantity		-0.028^a		-0.032^a		-0.028^a
		(0.003)		(0.004)		(0.006)
Pub*Quality		0.018^a		0.013^a		0.014^a
		(0.001)		(0.002)		(0.002)
A-rank in a national department						
Women	-0.165^a	-0.086^a	-0.164^a	-0.078^a	-0.165^a	-0.104^a
	(0.003)	(0.005)	(0.003)	(0.005)	(0.009)	(0.010)
Age		0.017^a		0.016^a		0.016^a
		(0.001)		(0.001)		(0.002)
Age^2		-0.000^a		-0.000^a		-0.000^b
		(0.000)		(0.000)		(0.000)
Publisher (Pub)		0.131^a		0.158^a		0.079^a
		(0.009)		(0.011)		(0.016)
Pub*Quantity		0.071^a		0.080^a		0.048^a
		(0.003)		(0.003)		(0.006)
Pub*Quality		0.003^b		0.009^a		-0.003
		(0.001)		(0.002)		(0.003)
A-rank in an international department						
Women	-0.082^a	0.023^a	-0.065^a	0.023^a	-0.139^a	0.020^c
	(0.003)	(0.004)	(0.003)	(0.005)	(0.010)	(0.011)
Age		0.001		-0.003^a		0.022^a
		(0.001)		(0.001)		(0.002)
Age^2		0.000^a		0.000^a		-0.000^a
		(0.000)		(0.000)		(0.000)
Publisher (Pub)		0.128^a		0.147^a		0.048^a
		(0.008)		(0.009)		(0.015)
Pub*Quantity		0.075^a		0.072^a		0.079^a
		(0.002)		(0.002)		(0.005)
Pub*Quality		0.026^a		0.028^a		0.024^a
		(0.001)		(0.001)		(0.002)
Specialisation	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	Yes	No	No	No	Yes
pseudo-R^2	0.022	0.255	0.021	0.228	0.031	0.396
Observations	37215	37215	30312	30312	6903	6903
log-likelihood	-46879	-35723	-37028	-29195	-9220	-5743

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

The marginal effects obtained from the multinomial regressions are reported in Table 4.9, while the coefficients are reported in Appendix 4.B. Looking at the entire sample (first two columns, with and without controls other than gender) we find a positive and significant effect of being female on the probability of being rank B in international departments and a negative one for being rank A in national ones. When we compare across the two types of affiliations, being a woman increases the probability of being rank B in a national department both for professors and researchers, with the effect being slightly larger for the latter (though not significantly different from that for professors). The probability of being rank A in a national department is reduced by being female, and the effect is substantially larger for researchers, contrary to our expectation. This evidence indicates that the salary-prestige trade-off is not the reason behind the gender gap.

Research output has the expected impact, reducing the probability of being rank B and increasing that of being rank A. The quantity of publications has a similar effect on the probability of being rank-A in either type of department, taking a value of 0.071 for national ones and of 0.075 for international ones, while quality has a much larger effect on the probability of being rank A in an international department (0.026 against 0.003). This comes as no surprise, implying that international departments are more selective.

The fourth panel presents puzzling results. Although there is a substantial raw gap in the probability of being rank A in an international department (being a woman decreases the probability by 8 percentage points), once we control for research output the effect becomes positive. The result holds for both university professors and researchers, although for the latter the effect is significant only at the 10 per cent level. This effect seems to be different from that obtained for rank-B in international departments, where we found a raw gap in favour of women that is substantially reduced once we include characteristics, implying that women in this category have a higher research output than men in this category. This effect is particularly large for researchers, where it falls from 0.147 to 0.064. In contrast, the results for rank-A in international departments indicate that women in this category have worse characteristics than men (either because they are younger or because of a lower output), in line with the results in Table 4.6. They are nevertheless more likely than similar men to be in an international department, and, as we said before, this could be due to positive discrimination, unobserved characteristics, or joint offers. With the data that we have at hand it is not possible to distinguish between these mechanisms.

To sum up, once we control for research output there is no difference across the genders in the probability of holding a junior position in a national department, while being a woman decreases the probability of being rank-A in those departments and increases that of being in an international department for either type of position. Although there seems to be a trade-off between being rank-B in a better department or rank-A in a less good one, the similarity between the coefficients for professors and researchers indicates that the cost of moving following a promotion is not the reason behind this. The case of researcher is particularly surprising, since it indicates that those in international departments and of B-rank have particularly high characteristics (relative to men), as witnessed by the drop in the coefficient on gender and yet remain unpromoted. It seems to us that the only possible explanation is the self-censorship that may occur in the male-dominated, highly competitive environment of a prestigious department that makes high ability women choose not to seek promotion.

4.4.3. Additional results

4.4.3.1. Network effects

One possible reason for differences in promotion probabilities is that men and women have different networks. The idea that networks are important in obtaining jobs and achieving promotions is widespread in the literature, and the issue has been addressed for promotions in academia; see McDowell and Smith (1992) for the US, Combes, Linnemer and Visser (2008) for France and Zinovyevay and Bagues (2012) for Spain. Coauthor networks have been shown to differ across genders, with females having fewer coauthors and a lower fraction of male coauthors; see McDowell and Smith (1992). If women have smaller or less efficient networks, then the probability of finding a member of the network in a promotion committee is lower than for a man, and this may affect the outcome.

In order to test this hypothesis, we construct for each individual three measures of networks. Our measures are based on coauthorship, obviously an imperfect measure of actual networks, but one that is quantifiable with the EconLit data. Hence, we consider the average number of authors per paper (*Authors number)*, the size of the network defined as the total number of different coauthors the researcher has had, and the fraction of network members that are men.

Table 4.10 reports logit regressions for promotions and affiliation, adding the three network characteristics as explanatory variables, as well as these interacted with gender. The effect of the average number of authors per article is unclear, with sign and significance of the estimated marginal effect changing across specifications. This could be due to the fact that although a bigger network has certain advantages, having more coauthors may

Table 4.10: Individual regressions with networks, panel 1990-2008, marginal effects

	Total sample		University		Researchers	
	(1) A-rank	(2) int.dept	(3) A-rank	(4) int.dept	(5) A-rank	(6) int.dept
Women	-0.109^a	0.073^a	-0.098^a	0.076^a	-0.097^a	0.077^a
	(0.006)	(0.007)	(0.007)	(0.008)	(0.017)	(0.016)
Age	0.015^a	-0.009^a	0.011^a	-0.010^a	0.037^a	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Age^2	0.000^b	0.000^a	0.000^a	0.000^a	-0.000^a	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Publisher (Pub)	0.210^a	0.177^a	0.252^a	0.190^a	0.074^a	0.066^a
	(0.009)	(0.011)	(0.011)	(0.013)	(0.021)	(0.021)
Pub*Quantity	0.133^a	0.038^a	0.131^a	0.045^a	0.108^a	0.010
	(0.004)	(0.005)	(0.004)	(0.006)	(0.011)	(0.010)
Pub*Quality	0.024^a	0.042^a	0.029^a	0.041^a	0.016^a	0.035^a
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Pub*Authors number	-0.016	0.005	-0.041^b	0.066^a	-0.023	-0.144^a
	(0.017)	(0.019)	(0.018)	(0.022)	(0.038)	(0.039)
Pub*Network size	0.042^a	0.048^a	0.054^a	0.031^a	0.052^a	0.090^a
	(0.007)	(0.008)	(0.008)	(0.009)	(0.015)	(0.015)
% Men in Network	0.054^a	-0.012	0.062^a	-0.039^a	0.072^a	0.079^a
	(0.009)	(0.010)	(0.009)	(0.012)	(0.020)	(0.021)
Women*Pub*Authors number	0.109^a	-0.017	0.075^a	-0.080^a	0.070	0.163^a
	(0.026)	(0.027)	(0.029)	(0.030)	(0.063)	(0.054)
Women*Pub*Network size	0.019^c	0.005	0.074^a	0.054^a	-0.054^b	-0.144^a
	(0.011)	(0.012)	(0.013)	(0.014)	(0.023)	(0.021)
Women*% Men in Network	-0.002	0.008	-0.012	0.010	0.075^c	0.055
	(0.017)	(0.018)	(0.018)	(0.020)	(0.041)	(0.038)
Specialisation	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	Yes	Yes	No	No	Yes	Yes
pseudo-R^2	0.370	0.127	0.396	0.068	0.419	0.368
Observations	37138	37138	30280	30280	6858	6858
log-likelihood	-15426	-20642	-11726	-17306	-2762	-2996

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

imply that promotion/recruitment committees see the individual as having done less work and 'discount' her/his research output, thus leading to the unstable results. The size of the networks has a positive effect on the probability of promotion and of being in an international department, while the fraction of men also affects positively the probability of promotion, though it has different effects on affiliation across our three populations. When we look at the interacted terms we find that the size of the network has a stronger effect than for men for female professors, though not for researchers. Despite the significance of these effects, the coefficients on gender are unchanged as compared to our earlier specifications. That is, gender gaps in networks do not seem to explain the lower promotion probabilities of women. Surprisingly, although the proportion of men in the network matters for the probability of promotion, it matters equally for males and females.

4.4.3.2. Changes over time

A number of recent papers have focussed on the evolution over time of gender gaps in promotions. For example, McDowell et al. (2001) find that by the 1980s there were no unexplained differences in promotion amongst academic economists in the US, although their results have subsequently been questioned; see, for example, Ginther and Kahn (2004). Admittedly, our data is not ideally suited to address this question since it spans only 19 years. We can nevertheless divide the sample into two subsamples to see if there are differences between the patterns observed in the 1990s and in the 2000s. The results are reported in Tables 4.11 and 4.12, with the former presenting estimations for the probability of promotion and the latter looking at affiliation to an international department. In both cases we report the results for the entire sample and those for university professors only, while omitting those for researchers which are very close to the reported results.

When we consider the determinants of promotion (Table 4.11), we find that the raw gap is not significantly different across subperiods, wether for the whole sample or for university professors only. However, once we introduce explanatory variables we obtain a large difference in the effect of gender, which almost halves across the two subperiods. The reason why we nevertheless observe no changes in the raw gap despite the decline in

Table 4.11: Probability to hold a A-rank position, marginal effects

	Total sample				University			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1990s	1990s	2000s	2000s	1990s	1990s	2000s	2000s
Women	-0.250^a	-0.088^a	-0.243^a	-0.051^a	-0.232^a	-0.078^a	-0.226^a	-0.040^a
	(0.007)	(0.008)	(0.006)	(0.006)	(0.007)	(0.008)	(0.006)	(0.007)
Age		0.019^a		0.015^a		0.016^a		0.009^a
		(0.002)		(0.001)		(0.002)		(0.001)
Age^2		-0.000		0.000		0.000		0.000^a
		(0.000)		(0.000)		(0.000)		(0.000)
Publisher (Pub)		0.293^a		0.250^a		0.342^a		0.282^a
		(0.012)		(0.012)		(0.015)		(0.014)
Pub*Quantity		0.140^a		0.152^a		0.148^a		0.155^a
		(0.004)		(0.003)		(0.005)		(0.004)
Pub*Quality		0.026^a		0.035^a		0.036^a		0.039^a
		(0.002)		(0.002)		(0.003)		(0.002)
Specialisation	No	Yes	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	Yes	No	Yes	No	No	No	No
pseudo-R^2	0.036	0.357	0.043	0.376	0.034	0.376	0.041	0.393
Observations	16935	16935	20280	20280	13622	13622	16690	16690
log-likelihood	-10924	-7286	-12630	-8236	-8593	-5553	-10087	-6387

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively.

the impact of gender is that the effect of the quantity and quality of publications increases between the 1990s and the 2000s and, since women have on average a lower research output, this offsets the direct effect of gender. The impact of age is stronger in the earlier period. Overall, the change in the estimated marginal effects seems to indicate that promotions have become more 'meritocratic' and less gender and age-related; However this has not resulted in an improvement in female's promotion rates since they tend to have a lower research output.

Table 4.12 reports the results for the probability of being in an international department. In this case the changes are substantial, and the raw gap is insignificant in the second decade. The effect of gender in the regression with all controls falls by about a quarter, from 0.086 to 0.061 in the case of the entire sample, and from 0.081 to 0.067 for

Table 4.12: Probability to be in an international department, marginal effects

	Total sample				University			
	(1) 1990s	(2) 1990s	(3) 2000s	(4) 2000s	(5) 1990s	(6) 1990s	(7) 2000s	(8) 2000s
Women	0.044[a]	0.086[a]	0.002	0.061[a]	0.037[a]	0.081[a]	0.024[a]	0.067[a]
	(0.009)	(0.009)	(0.007)	(0.008)	(0.010)	(0.010)	(0.008)	(0.009)
Age		-0.005[a]		-0.010[a]		-0.008[a]		-0.011[a]
		(0.002)		(0.001)		(0.002)		(0.002)
Age^2		0.000[a]		0.000[a]		0.000[a]		0.000[a]
		(0.000)		(0.000)		(0.000)		(0.000)
Publisher (Pub)		0.205[a]		0.183[a]		0.211[a]		0.197[a]
		(0.013)		(0.014)		(0.016)		(0.016)
Pub*Quantity		0.065[a]		0.050[a]		0.062[a]		0.048[a]
		(0.005)		(0.004)		(0.006)		(0.005)
Pub*Quality		0.044[a]		0.050[a]		0.041[a]		0.049[a]
		(0.002)		(0.002)		(0.003)		(0.002)
Specialisation	No	Yes	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	Yes	No	Yes	No	No	No	No
pseudo-R^2	0.001	0.107	0.000	0.136	0.001	0.067	0.001	0.068
Observations	16935	16935	20280	20280	13622	13622	16690	16690
log-likelihood	-10700	-9563	-12966	-11203	-8369	-7822	-10193	-9506

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively.

university professors. The effect of research output has also changed over time, with the effects of quantity falling and those of quality increasing between the first and the second decade. All this indicates that the 'advantage' that women have at being in an international department has diminished over time. A tentative conclusion is that, given that the practice of joint appointments has increased in the past decade or so, this is unlikely to be the main mechanism behind the higher presence of women in international departments.

4.5. Conclusion

In this paper we have examined the promotion gap amongst academic economists in France. The novelty of our analysis is twofold. First, although a substantial literature on promotion gaps in academia exists, it has focused on the US and the UK, and there is little evidence concerning the large continental European economies. The interest of the exercise lies in that the Anglo-Saxon and the European academic labour markets are substantially different. The former involve substantial initial risks (before tenure is awarded), rely on an internal decision-making process for promotion, exhibit flexible salary scales, and display a trade-off between the prestige of the department and promotion thresholds. In contrast, in the French system academics are public servants that have tenured positions from the moment they are first recruited, salaries are fixed by a common grid, and promotions occur through a national contest taking place outside the candidate's department. Despite these differences, the effect of gender that we obtain is remarkably close to that found for US academics. Ginther and Kahn (2004) find a raw gender gap of -0.213, which falls to -0.130 once age and publication records are included. In our data, the respective values of the gender gap are -0.246 and -0.124.

The second contribution of our paper is to exploit the specificities of the French academic system to test for different potential explanations for the gender promotion gap. This gap could be due to either discrimination or women's behaviour. We test for discrimination in two ways. First, we compare the effect of being a woman on the probability of being promoted and on the probability of being in a prestigious department. Our findings indicate that, once we control for research output, women are more likely to be in a prestigious department than men. Since promotion and hiring decisions are typically made by the same individuals, it is hard to see why they would discriminate against women in one dimension and not in another. Second, we allow for differences in the returns to characteristics across genders and although the returns to experience are lower for women, in line with a large literature, the returns to research output are higher than for men. Again, such evidence is difficult to reconcile with any form of discrimination.

We then explore two possible behavioural differences: that promotion is more costly for women than for men and hence they are less likely to seek promotion, and that women have a stronger preference for department prestige as compared to salary than men, which makes them choose to stay in prestigious departments at the cost of not being promoted. We exploit the fact that there are two types of academic positions in France, one of which requires changing department after promotion and one which does not, to test for these two hypotheses. The empirical results lend no support to either of them.

Our analysis then raises the question of why women experience lower promotion probabilities. A possible answer is that women are less willing to enter into contests or that they perform below their potential ability in contests against men, both of which would reduce the observed frequency of promotions for women. Evidently, the policy implications of such an explanation are very different from those stemming from discrimination or differences in costs and preferences. For example, major orchestras have traditionally been perceived as discriminating strongly against women, and the introduction of blind auditions seems to have enhanced substantially both female hiring and promotion. In the case of female academic economists, discrimination does not seem to be the cause of the gender gap and antidiscriminatory policies of this type are unlikely to work. Encouraging women to seek promotion seems to be a more suitable approach, although it is unclear how to foster it in the short-term. One possibility would be to establish a system of mentoring along the lines of that proposed by the Committee on the Status of Women in the Economics Profession of the American Economic Association.

APPENDIX

4.A. Definition of international departments

Academia in France is organised around 'research centers', with a university potentially having several research centers in Economics. We have defined “departments” either by an economics department when it is the single affiliation where economists are found in a given university (which corresponds to the majority of cases), or by the aggregation of all research centres where there are economists in the university. We performed robustness checks by looking at research centres rather than departments, and obtained consistent results (available upon request). However, our notion of slightly aggregated economics departments better matches the reality of French academic research and hence have preferred to focus our discussion of the results on this concept.

We define “international departments" as departments with the highest research output, measured by both the total stock and average stock of publications per member of the department in EconLit journals. Also, we wish to take into account both “normal" quality (CLm) and top quality (CLh) research; see Bosquet and Combes (2012). In order to do so, we apply the following procedure at each point in time. We first calculate the total research outputs of a department (in CLm and CLh) as the sum of research outputs of its members. We also calculate the two scores per member of the department. Hence, we obtain 4 scores per department: CLm total, CLh total, average CLm and average CLh. Departments are then ranked with respect to these 4 scores and their final score is the average of these rankings.

The cutoff point for top departments is arbitrary. We choose our cutoff at each point in time so as to have 35% of the total population in the top department category. This choice implies that the same fraction of the population in the sample is in a top department as has a rank A job. Since departments may change their status over time, we considered as international departments for the entire sample period those departments which are more often than not in the top department category.

Table 4.13: List of international departments

cepremap	1990	2004
crest-ensae	1993	2008
ec. polytechnique	1990	2008
ehess-ens	1990	2004
enpc	1990	2004
hec	1995	2008
u. aix marseille 2-3	1990	2008
u. cergy pontoise	2005	2008
u. cergy pontoise-paris 10	1990	2004
u. clermont 1	1990	2008
u. nancy 2-strasbourg 1	2006	2008
u. paris 1	1990	2004
u. paris 1-eep	2005	2008
u. paris 10	2005	2008
u. paris 9	1990	2008
u. toulouse 1-inra	1990	2008

4.B. Coefficients of the multinomial regressions

Table 4.14: Multinomial regressions, panel 1990-2008

	Total sample		University		Researchers	
	(1)	(2)	(3)	(4)	(5)	(6)
B-rank in an international department						
Women	0.202^a	0.298^a	0.220^a	0.300^a	0.139^c	0.425^a
	(0.031)	(0.033)	(0.034)	(0.036)	(0.075)	(0.096)
Age		-0.051^a		-0.048^a		-0.060^a
		(0.007)		(0.008)		(0.019)
Age^2		0.002^a		0.002^a		0.001^b
		(0.000)		(0.000)		(0.000)
Publisher (Pub)		1.154^a		1.133^a		0.853^a
		(0.056)		(0.065)		(0.136)
Pub*Quantity		0.169^a		0.123^a		0.242^a
		(0.026)		(0.029)		(0.062)
Pub*Quality		0.235^a		0.198^a		0.270^a
		(0.012)		(0.014)		(0.026)
Constant	-0.862^a	-0.699^a	-0.956^a	-0.733^a	-0.456^b	-0.541^c
	(0.078)	(0.107)	(0.088)	(0.118)	(0.178)	(0.284)
A-rank in a national department						
Women	-1.309^a	-0.614^a	-1.290^a	-0.536^a	-1.385^a	-0.836^a
	(0.040)	(0.046)	(0.044)	(0.051)	(0.097)	(0.114)
Age		0.141^a		0.114^a		0.227^a
		(0.009)		(0.010)		(0.022)
Age^2		-0.000		0.000^b		-0.002^a
		(0.000)		(0.000)		(0.000)
Publisher (Pub)		2.028^a		2.363^a		1.237^a
		(0.066)		(0.075)		(0.148)
Pub*Quantity		1.040^a		1.126^a		0.865^a
		(0.026)		(0.030)		(0.062)
Pub*Quality		0.226^a		0.283^a		0.156^a
		(0.013)		(0.015)		(0.028)
Constant	-0.272^a	-4.115^a	-0.413^a	-3.993^a	0.282^c	17.835
	(0.070)	(0.132)	(0.079)	(0.148)	(0.156)	(.)
A-rank in an international department						
Women	-0.945^a	0.032	-0.864^a	0.053	-1.117^a	0.051
	(0.041)	(0.054)	(0.048)	(0.062)	(0.086)	(0.128)
Age		0.079^a		0.022^c		0.325^a
		(0.010)		(0.012)		(0.027)
Age^2		0.001^a		0.003^a		-0.004^a
		(0.000)		(0.000)		(0.001)
Publisher (Pub)		2.853^a		3.358^a		1.313^a
		(0.083)		(0.097)		(0.180)
Pub*Quantity		1.476^a		1.546^a		1.322^a
		(0.031)		(0.035)		(0.075)
Pub*Quality		0.474^a		0.533^a		0.419^a
		(0.014)		(0.016)		(0.030)
Constant	-0.837^a	-4.786^a	-0.947^a	-4.440^a	-0.387^b	18.980^a
	(0.083)	(0.162)	(0.094)	(0.182)	(0.185)	(0.433)
Specialisation	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Belonging	No	Yes	No	No	No	Yes

continued on next page

continued from previous page

	Total sample		University		Researchers	
	(1)	(2)	(3)	(4)	(5)	(6)
pseudo-R^2	0.022	0.255	0.021	0.228	0.031	0.396
Observations	37215	37215	30312	30312	6903	6903
log-likelihood	-46879	-35723	-37028	-29195	-9220	-5743

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively.

4.C. Difference in differences

Table 4.15: Diff. in diff. regressions, panel 1990-2008, marginal effects

	Logit		Probit	
	(1) A-rank	(2) int.dept	(3) A-rank	(4) int.dept
Women	-0.066[a]	0.077[a]	-0.057[a]	0.075[a]
	(0.011)	(0.015)	(0.010)	(0.015)
Women*University	-0.002	-0.006	-0.011	-0.005
	(0.012)	(0.015)	(0.012)	(0.015)
Age	0.057[a]	-0.004[c]	0.055[a]	-0.005[b]
	(0.002)	(0.003)	(0.002)	(0.002)
Age^2	-0.001[a]	0.000	-0.001[a]	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Age*University	-0.050[a]	-0.003	-0.048[a]	-0.003
	(0.002)	(0.003)	(0.002)	(0.003)
Age^2*University	0.001[a]	0.000[a]	0.001[a]	0.000[a]
	(0.000)	(0.000)	(0.000)	(0.000)
Publisher (Pub)	0.214[a]	0.170[a]	0.203[a]	0.172[a]
	(0.016)	(0.019)	(0.015)	(0.019)
Pub*Quantity	0.111[a]	0.075[a]	0.110[a]	0.074[a]
	(0.006)	(0.007)	(0.006)	(0.007)
Pub*Quality	0.026[a]	0.054[a]	0.025[a]	0.054[a]
	(0.002)	(0.003)	(0.002)	(0.003)
Publisher (Pub)*University	0.082[a]	0.024	0.091[a]	0.027
	(0.015)	(0.019)	(0.015)	(0.019)
Pub*Quantity*University	0.045[a]	-0.027[a]	0.048[a]	-0.026[a]
	(0.007)	(0.008)	(0.007)	(0.008)
Pub*Quality*University	0.010[a]	-0.010[a]	0.012[a]	-0.009[b]
	(0.003)	(0.004)	(0.003)	(0.004)
Specialisation	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Belonging	Yes	Yes	Yes	Yes
pseudo-R^2	0.383	0.129	0.383	0.129
Observations	37215	37215	37215	37215
log-likelihood	-15142	-20630	-15162	-20625

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively.

Part III

Econometrics Issues when Estimating International Trade Gravity Equations

Chapter 5

Comments on "The Log of Gravity"[1]

Relying on the idea that OLS estimates of log-linearised models are biased, more and more papers have started to use non-linear estimators in levels for gravity equations, following the seminal work of Santos Silva and Tenreyro (2006). I discuss here the hypothesis on which this analysis is based and explain why this could be the case that the non-linear estimators in levels are biased when log-linear OLS are not. I present a specification test for non-nested alternative hypotheses to discriminate between the two. Sadly, this test is vey often non-conclusive and it can only be implemented to discriminate between log-linear OLS and non-linear least squares estimates in levels. Evidences are supported by Monte Carlo simulations.

5.1. Introduction

In a very influential paper, Santos Silva and Tenreyro (2006) argue that the parameters of a log-linearised model estimated by ordinary least squares (OLS) are biased as a

1. I am particulary grateful to Thierry Magnac who provided the insight for this chapter.

consequence of Jensen's inequality, under heteroskedasticity in levels. As a consequence, they propose to use a non-linear estimator of such models directly in levels and argue that the Poisson Pseudo-Maximum Likelihood (PPML) estimator should be the most efficient one, especially for gravity equations. This last choice has been questioned by other simulation studies, including Martin and Pham (2008), Martìnez-Zarzoso, Nowak-Lehmann D. and Vollmer (2009) and Xiong and Chen (2012), but the crucial assumption on which all this research is based has not. Indeed, to obtain biased estimates from log-linear OLS, Santos Silva and Tenreyro (2006) assume a specification of the residuals in levels. Then, if the variance of these residuals depends on some explicative variables (heteroskedasticity), by taking logs, the log-linearised residuals are likely to be correlated with these explicative variables since the logarithm of a random variable does not only depend on its expectation but also on higher moments, including its variance. This violates the condition under which the OLS estimator is consistent (independence of the residuals with respect to the explicative variables) and parameters of interest are biased.

Nevertheless, there is, to my knowledge, no more *a priori* econometric reason to specify the residuals of an economic model in levels rather than in logs. The fact that taking logs create a selection issue when the dependent variable in levels is zero could be an argument. On the contrary, we could also argue that numerical methods are more stable in linear applications and that if the selection process follows a different pattern than positive values, then simply keeping the zeros in the sample will not be an appropriate solution. As a consequence, if the true underlying data generating process (including residuals) is in levels, then the Santos Silva and Tenreyro (2006)'s methodology should be applied, but if the true underlying data generating process is log-linear, then the non-linear estimates in levels are going to be biased. Indeed, I first show in this paper that the reverse of Santos Silva and Tenreyro (2006)'s proposition is also true, mathematically and with Monte Carlo simulations.

From this point, if there is no *a priori* reason to prefer a specification in levels rather than in logs, it could be useful to use some testing to discriminate between the two alternatives in an applied context. In order to do so, I propose to use a test for model specification in the presence of non-nested alternative hypotheses. This builds on MacKinnon, White and Davidson (1983) who extend in several directions the result established by Davidson and MacKinnon (1981), in particular to the case where the two non-nested models involve different transformations of the dependent variable, which is the case of interest here. Sadly, this test can only be used with normally and identically distributed residuals and, hence, allows only to try to discriminate between log-linear OLS and non-linear least squares in levels. In particular, this rules out the possibility to try to discriminate between log-linear OLS and the estimates from the PPML estimator in levels proposed by Santos Silva and Tenreyro (2006).

In this paper, several sets of Monte Carlo simulations illustrate the following results. First, I confirm the results exposed by Santos Silva and Tenreyro (2006): in the case of a data generating process in levels, OLS estimates of the log-linearised counterpart are biased and the PPML estimator in levels performs well, even in the cases where the non-linear least squares (NLS) and the gamma pseudo-maximum likelihood (GPML) estimators are optimal. Second, I show that the reverse of this proposition is also true by simulating a log-linear data generating process and by using the same non-linear estimator in levels (PPML, NLS, GPML) which lead to biased estimates of the parameters of interest. This bias is particularly important in the presence of heteroskedasticity. Third, I confirm that the specification test for non-nested alternative hypotheses of MacKinnon et al. (1983) can only be used to try to discriminate between log-linear OLS and non-linear least squares in levels by showing that the test is very often misleading or non-conclusive when the data generating process makes PPML or GPML optimal. Indeed, in these cases, both specifications (in levels and in logs) are simultaneously rejected or simultaneously accepted by the other one for most simulated samples.

Finally, I try to use the MacKinnon et al. (1983)'s test in the applied case of international trade gravity equations. Using the data and the specifications from Santos Silva and Tenreyro (2006), the test is non-conclusive and does not allow to choose between non-linear estimates in levels and the OLS estimates of the log-linearised counterpart.

The rest of the paper is organised as follows. Section 5.2 discusses the heteroskedasticity issue introduced by Santos Silva and Tenreyro (2006) and presents the reversibility of the argument. Section 5.3 illustrates this result with Monte Carlo simulations. Section 5.4 presents how MacKinnon et al. (1983)'s P_ϵ test can be adapted to choose between these alternative specifications and discusses the conditions under which the test performs well using Monte Carlo simulations. Section 5.5 uses this test to try to discriminate between log-linear OLS and non-linear least squares estimates in levels of the gravity equations and section 5.6 concludes.

5.2. Illustration of the heteroskedasticity issue

The argument developed by Santos Silva and Tenreyro (2006) can be summarised as follows. Lets y_i be the dependent variable, x_i the vector of explicative variables and β

the associated vector of parameters and lets assume that the true underlying model is $y_i = exp(x_i\beta)\eta_i$ in its stochastic version, with an error term η_i statistically independent of the regressors: $E(\eta_i|x_i) = 1$ and $E(y_i|x_i) = exp(x_i\beta)$. Then, the common practice of log-linearisation consists in estimating $log\ y_i = x_i\beta + log\ \eta_i$ by OLS. Santos Silva and Tenreyro (2006) highlight that this procedure is valid only if the new error term $(log\ \eta_i)$ is statistically independent of the regressors. The problem arises in case of heteroskedasticity in levels, this is if the variance of the error term η_i depends on x_i, because the logarithm of a random variable depends both on its mean and on higher moments including the variance. Hence, if $V(\eta_i|x_i)$ depends on x_i, it is likely that $E(log\ \eta_i|x_i)$ also depends on x_i as a consequence of Jensen's inequality $(E(log\ Z) \neq log\ E(Z))$ and this violates the condition of consistency of the OLS estimator.

However, I argue here that the reverse is also true. Indeed, if the true underlying model is $log\ y_i = x_i\ \gamma + log\ \epsilon_i$ with an error term statistically independent of the regressors, $E(log\ \epsilon_i|x_i) = 0$ and $E(log\ y_i|x_i) = x_i\ \gamma$; then estimating $y_i = exp(x_i\ \gamma)\epsilon_i$ will be biased in the case of heteroskedasticity. If the variance of $log\ \epsilon_i$ depends on x_i then it is likely that $E(\epsilon_i|x_i)$ also depends on x_i as another consequence of Jensen's inequality $(E[exp(Z)] \neq exp(E[Z]))$. Next section illustrates these results with Monte Carlo simulations.

5.3. Monte Carlo simulations

This section illustrates the results presented in section 5.2. To focus on what is at stake, I build the most possible simple specification with only one explicative variable and by getting rid of the selection issue. Augmented specifications with one additional explicative variable are presented in Appendix 5.A. The added variable is a dummy variable which equals one with probability 0.4, which yields exactly the same simulation scheme as Santos Silva and Tenreyro (2006). It is unnecessary for the argument made here, simulations including it lead to the same conclusions.

All presented simulations are based on the following assumptions:

$$X \sim \mathcal{N}(0,1) \tag{5.1}$$

$$\beta_0 = 0 \qquad \text{and} \qquad \beta_1 = 1 \tag{5.2}$$

I use a sample size of 1,000.

5.3.1. Specification in levels

The first set of simulations re-illustrates Santos Silva and Tenreyro (2006)'s argument. Is is based on the additional following assumptions.

$$E[y_i|x_i] = \mu(x_i\beta) = exp(\beta_0 + \beta_1 x_i) \tag{5.3}$$

$$y_i = \mu(x_i\beta)\eta_i \tag{5.4}$$

where η_i is a log normal random variable such that $E[\eta_i] = 1$ and $V(\eta_i) = \sigma_i^2$ given by:

$$\text{Case 1: } \sigma_i^2 = \mu(x_i\beta)^{-2} \longrightarrow \text{NLS optimal} \tag{5.5}$$

$$\text{Case 2: } \sigma_i^2 = \mu(x_i\beta)^{-1} \longrightarrow \text{PPML optimal} \tag{5.6}$$

$$\text{Case 3: } \sigma_i^2 = 1 \longrightarrow \text{GPML optimal} \tag{5.7}$$

Case 1 is the standard case where the non-linear least squares (NLS) estimator is optimal. Indeed, $V(y_i|x_i) = V(\mu(x_i\beta)\eta_i|x_i) = \mu(x_i\beta)^2\ V(\eta_i|x_i) = \mu(x_i\beta)^2\ \mu(x_i\beta)^{-2} = 1$. Case 2 is the standard case where the Poisson Pseudo-Maximum Likelihood (PPML) estimator is optimal: $V(y_i|x_i) = \mu(x_i\beta)^2\ V(\eta_i|x_i) = \mu(x_i\beta)^2\ \mu(x_i\beta)^{-1} = \mu(x_i\beta) = E[y_i|x_i]$. Case 3 is the standard case where the gamma Pseudo-Maximum Likelihood (GPML) estimator is optimal: $V(y_i|x_i) = \mu(x_i\beta)^2 = E[y_i|x_i]^2$.

On this simulation setting, we compare the performances of the NLS, PPML and GPML estimators, which directly estimate equation 5.4, and of the OLS estimator of its log-linearised counterpart: $log\ y_i = x_i\beta + log\ \eta_i$.

Table 5.1: Monte Carlo simulations, 10,000 replicas

	β_0		β_1	
Estimator	Bias	S.E.	Bias	S.E.
	Case 1: $V[y_i|x] = 1 \longrightarrow$ NLS optimal			
PPML	-0.00066	0.04283	0.00023	0.02816
GPML	-0.00583	0.06121	0.01506	0.08608
NLS	0.00001	0.02802	-0.00004	0.01348
OLS	-0.53388	0.03393	0.50069	0.04369
	Case 2: $V[y_i|x] = \mu(x_i\beta) \longrightarrow$ PPML optimal			
PPML	-0.00060	0.03524	-0.00011	0.02496
GPML	-0.00189	0.03955	0.00392	0.04980
NLS	-0.00245	0.06412	0.00039	0.04283
OLS	-0.40282	0.02847	0.24976	0.03253
	Case 3: $V[y_i|x] = \mu(x_i\beta)^2 \longrightarrow$ GPML optimal			
PPML	-0.00115	0.04938	-0.00230	0.06729
GPML	-0.00107	0.03193	0.00014	0.03156
NLS	-0.31900	7.65695	0.09594	2.57697
OLS	-0.34656	0.02669	0.00017	0.02620

S.E. = Standard Error. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

Table 5.1 presents the results based on 10,000 replicas, for the average bias of the parameters of interest β_0 and β_1 with their standard errors. As expected, all estimators in levels are consistent in the three cases. Each one of them outperfoms the two others in the case where it is optimal. OLS estimates are biased, except for β_1 when GPML is optimal. This corresponds to the case of homoskedasticity in levels for which OLS estimation of the log-linear model is consistent for the slope parameters, as argued by Santos Silva and Tenreyro (2006). PPML also performs relatively well in the cases where NLS and GPML are optimal.

5.3.2. Specifications in logs

The second set of simulations illustrates the reciprocity of Santos Silva and Tenreyro (2006)'s argument. It is based on the additional following assumptions:

$$E[log\ y_i|x] = x_i\beta = \beta_0 + \beta_1 x_i \tag{5.8}$$

$$log\ y_i = x_i\beta + \epsilon_i \tag{5.9}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ is given by:

$$\text{Case 4: } \sigma_i^2 = 1 \text{ (homoskedastic)} \tag{5.10}$$

$$\text{Case 5: } \sigma_i^2 = \mu(x_i\beta) \text{ (heteroskedastic)} \tag{5.11}$$

In case 4, the variance of the residuals is constant, this is the standard case of homoskedasticity and in case 5, I introduce a particular form of heteroskedasticity where the variance of the residuals is proportional to $exp(x_i\beta)$.

On this simulation setting, we compare the performances of the OLS estimator which directly estimates equation 5.9 and of the NLS, PPML and GPML estimators which estimate $y_i = exp(\beta_0 + \beta_1 x_i + \epsilon_i)$.

Table 5.2: Monte Carlo simulations, 10,000 replicas

	β_0		β_1	
Estimator	Bias	S.E.	Bias	S.E.
	Case 4: $\sigma_i^2 = 1$ (homoskedastic)			
PPML	0.49903	0.06336	-0.00465	0.08336
GPML	0.49844	0.04197	0.00016	0.04134
NLS	0.05844	7.34495	0.13068	2.47526
OLS	-0.00014	0.03214	0.00043	0.03196
	Case 5: $\sigma_i^2 = \mu(x_i\beta)$ (heteroskedastic)			
PPML	-1.48398	22.44447	2.02282	7.12034
GPML	0.94775	0.32411	0.80008	0.28027
NLS	-27.94363	419.79984	10.66391	147.60112
OLS	-0.00017	0.04074	-0.00042	0.05792

S.E. = Standard Error. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

Here also, Table 5.2 presents the results based on 10,000 replicas, for the average bias of the parameters of interest β_0 and β_1 with their standard errors. As expected, the OLS estimator is consistent under both cases and a little less efficient (larger variance) where there is some heteroskedasticity. NLS, PPML and GPML estimates are biased, except for GPML estimate of β_1 in the homoskedastic case. Bias are much larger in case of heteroskedasticity.

5.4. Specification testing

Given, first, that OLS estimates are biased if the data generating process is in levels, second, that non-linear estimates are biased if the data generating process is log-linear and third, that we have no prior on the data generating process in applied cases, it could be very useful to find a way to discriminate between log-linear OLS and non-linear estimates in levels. This is the purpose of this section.

5.4.1. The P_ϵ test

In order to do so, we can use the P_ϵ (extended P) specification test for non-nested alternative hypotheses of MacKinnon et al. (1983). This is an extension of Davidson and MacKinnon (1981)'s J and P test, in particular to the case where the two models involve different transformations of the dependent variable.

The test is stated as follows. The null and alternative hypotheses to be tested are:

$$\begin{aligned} H_0 &: y_i = f(X_i, \beta) + \epsilon_{0i} \\ H_1 &: h(y_i) = g(Z_i, \gamma) + \epsilon_{1i} \end{aligned} \tag{5.12}$$

where $h(.)$ may be any monotonic, continuously differentiable function which does not depend on any unknown parameters. ϵ_{0i} and ϵ_{1i} are normally and identically distributed $\mathcal{N}(0, \sigma_j^2), j = 1, 2$.

This yields the artificial compound model:

$$H_c : (1-\alpha)\big(y_i - f(X_i, \beta)\big) + \alpha\big(h(y_i) - g(Z_i, \gamma)\big) = \epsilon_i \tag{5.13}$$

For simplicity, MacKinnon et al. (1983) propose to estimate the following regression:

$$y_i - f(X_i, \hat{\beta}) = \alpha\Big(g(Z_i, \hat{\gamma}) - h\big(f(X_i, \hat{\beta})\big)\Big) + \hat{F}\ b + \epsilon_i \tag{5.14}$$

where $\hat{\beta}$ and $\hat{\gamma}$ are the estimates of β and γ under the null and the alternative hypotheses, respectively. $\hat{F}$ is the row vector containing the derivatives of $f(X_i, \beta)$ with respect to β at the point estimate. If H_0 is true, then α is supposed to be not significant.

Empirically, it is possible that no specification appears to dominate the other. Indeed, in order to be conclusive, the test should be consistent when the two hypotheses are inverted: if H_0 is true, then α should be not significant when estimating equation 5.14 but it should also be simultaneously significant when estimating the transformation of equation 5.14 when H_1 and H_0 are the null and the alternative hypotheses, respectively.

The purpose of next section is to see under which conditions the test is conclusive in the case where we want to discriminate between a non-linear estimator in levels and its log-linearised counterpart.

5.4.2. Monte Carlo simulations

Following the analysis developed in sections 5.2 and 5.3, the null and alternative hypotheses that we are interested to test are:

$$\begin{aligned} &H_0 : y_i = exp(X_i\beta) + \epsilon_{0i} \\ &H_1 : ln\ y_i = X_i\ \gamma + \epsilon_{1i} \end{aligned} \tag{5.15}$$

which yields to test the value of α in:

$$y_i - exp(X_i\ \hat{\beta}) = \alpha\Big(X_i\ \hat{\gamma} - ln\big(exp(X_i\ \hat{\beta})\big)\Big) + X_i\ exp(X_i\ \hat{\beta})\ b + \epsilon_i \tag{5.16}$$

Alternatively, we also have to test for symmetry purpose:

$$H_0 : z_i = X_i\beta + \epsilon_{0i}$$
$$H_1 : exp(z_i) = exp(X_i\ \gamma) + \epsilon_{1i} \tag{5.17}$$

with $z_i = log\ y_i$, yielding to test the value of α in:

$$ln\ y_i - exp(X_i\ \hat{\beta}) = \alpha\Big(exp(X_i\ \hat{\gamma}) - exp(X_i\ \hat{\beta})\Big) + X_i\ b + \epsilon_i \tag{5.18}$$

Hence, if the true data generating process is non-linear, *i.e.* if H_0 is true in 5.15 and H_1 is true in 5.17, then α should be not significant in equation 5.16 but significant in equation 5.18. On the contrary, if the data generating process is log-linear, *i.e.* if H_1 is true in 5.15 and H_0 is true in 5.17, then α should significant in equation 5.16 but not significant in equation 5.18.

I use the same Monte Carlo simulations presented in section 5.3 to check under which conditions this testing procedure could be useful. In order to do so, I keep the same non-linear (cases 1, 2 and 3) and log-linear (cases 4 and 5) data generating processes also presented in section 5.3 and always test the three non-linear estimators in levels (NLS, PPML and GPML) against the log-linear OLS estimator.

For each case, Table 5.3 gives the percentage of the 10,000 replicas for which the test is valid, *i.e.* it discriminates in the right direction, and the percentage of the 10,000 replicas for which it is non-conclusive, *i.e.* it cannot discriminate in either direction (a non-linear estimator in levels or the log-linear OLS estimator; because both reject or both accept the

Table 5.3: Test levels VS logs, 10,000 replicas

	1%		5%		10%	
Estimator	% Ok	% N.C.	% Ok	% N.C.	% Ok	% N.C.
			Case 1: $V[y_i\|x] = 1 \longrightarrow$ NLS optimal			
PPML	0.950	0.049	0.970	0.031	0.933	0.067
GPML	0.934	0.063	0.952	0.048	0.925	0.076
NLS	0.951	0.048	0.970	0.030	0.934	0.066
			Case 2: $V[y_i\|x] = \mu(x_i\beta) \longrightarrow$ PPML optimal			
PPML	0.027	0.943	0.176	0.749	0.330	0.569
GPML	0.012	0.952	0.133	0.782	0.299	0.586
NLS	0.029	0.941	0.178	0.747	0.331	0.567
			Case 3: $V[y_i\|x] = \mu(x_i\beta)^2 \longrightarrow$ GPML optimal			
PPML	0.006	0.658	0.017	0.536	0.031	0.469
GPML	0.004	0.669	0.015	0.556	0.028	0.498
NLS	0.005	0.771	0.018	0.656	0.036	0.590
			Case 4: $\sigma_i^2 = 1$ (homoskedastic)			
PPML	0.302	0.694	0.423	0.559	0.472	0.495
GPML	0.283	0.713	0.404	0.579	0.451	0.516
NLS	0.197	0.799	0.317	0.661	0.381	0.575
			Case 5: $\sigma_i^2 = \mu(x_i\beta)$ (heteroskedastic)			
PPML	0.233	0.487	0.256	0.500	0.250	0.534
GPML	0.341	0.508	0.321	0.548	0.293	0.591
NLS	0.113	0.517	0.215	0.474	0.250	0.503

PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares. N.C. = Non Conclusive.

other specification). The complement is the percentage of the 10,000 replicas for which the test discriminates in the wrong direction. For each case, the log-linear OLS estimator is tested against the NLS, PPML and GPML estimators in levels and these percentages are given for several significance thresholds for α.

As expected, when the residuals are assumed to be normally and identically distributed whereas they are not (cases 2, 3 and 5), the test is very often non-conclusive and, more importantly, it also often concludes in the wrong direction. Hence, this test should be applied only in a context where the residuals are normally and identically distributed, which rules out the possibility to try to discriminate between the log-linear OLS estimator and the PPML estimator in levels proposed by Santos Silva and Tenreyro (2006). On the contrary, when the NLS estimator is optimal (case 1), the test is valid for more than 92% of the samples, with very few conclusions in the wrong direction (the remaining samples yielding non-conclusive testings). And when the data generating process is log-linear without heteroskedasticity, the test concludes also very rarely in the wrong direction even if the percentage of samples for which the test concludes in the right direction is never greater than 47%.

5.5. Application to the gravity equation

Following Santos Silva and Tenreyro (2006) and because their proposed method has been mainly applied to gravity equations, I try to use the P_ϵ test of MacKinnon et al. (1983) to discriminate between log-linear OLS and non-linear gravity equation estimates in levels. The data and the specifications (choice of variables) are taken from Santos Silva and Tenreyro (2006).

Results for the Anderson and van Wincoop (2003)'s gravity equation are presented here whereas results for the traditional gravity equation are given in Appendix 5.B. In both cases, I add the GPML estimates to Santos Silva and Tenreyro (2006)'s tables.

Table 5.4 shows that including or not zero trade flows only matters for GPML estimates. Point estimates substantially vary between the different estimators. In particular, the trade elasticity to distance is twice larger with OLS than with NLS in levels.

Table 5.5 shows the estimates and standard errors of α when it is estimated from

Table 5.4: The Anderson-van Wincoop gravity equation

	OLS	NLS > 0	NLS	PPML > 0	PPML	GPML > 0	GPML
Log distance	-1.347[a]	-0.591[a]	-0.582[a]	-0.770[a]	-0.750[a]	-1.173[a]	-1.933[a]
	(0.031)	(0.134)	(0.130)	(0.042)	(0.041)	(0.029)	(0.055)
Contiguity dummy	0.174	0.453[b]	0.458[b]	0.352[a]	0.370[a]	0.326[a]	-0.457[b]
	(0.130)	(0.203)	(0.205)	(0.090)	(0.091)	(0.117)	(0.230)
Common-language dummy	0.406[a]	0.937[a]	0.925[a]	0.418[a]	0.383[a]	0.416[a]	0.681[a]
	(0.068)	(0.187)	(0.187)	(0.094)	(0.093)	(0.065)	(0.097)
Colonial-tie dummy	0.666[a]	-0.748[b]	-0.736[b]	0.038	0.079	0.510[a]	0.808[a]
	(0.070)	(0.301)	(0.305)	(0.134)	(0.134)	(0.067)	(0.100)
Free-trade agreement dummy	0.310[a]	1.006[a]	1.017[a]	0.374[a]	0.376[a]	0.579[a]	1.472[a]
	(0.098)	(0.235)	(0.227)	(0.076)	(0.077)	(0.085)	(0.257)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9613	9613	18360	9613	18360	9613	18360

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

Table 5.5: Test on the Anderson-van Wincoop gravity equation

	without zeros		with zeros	
	(1)	(2)	(3)	(4)
	H_0: NLS	H_0: OLS	H_0: NLS	H_0: OLS
	H_1: OLS	H_1: NLS	H_1: OLS	H_1: NLS
Alpha	34425.409[a]	0.000[a]	35222.085[a]	0.000[a]
	(3836.587)	(0.000)	(3871.168)	(0.000)
Derivatives	Yes	Yes	Yes	Yes
Observations	9613	9613	9613	9613

Standard error between brackets. [a] Significant at the 1% level. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

equation 5.16 (columns (1) and (3)) and equation 5.18 (columns (2) and (4)) to try to discriminate between log-linear OLS (first column of Table 5.4) and non-linear least squares estimates in levels (second and third columns of Table 5.4). NLS estimates include (columns (1) and (2)) or not (columns (3) and (4)) zero trade flows. α is always significant at the 1% level, meaning that the null hypothesis H_0 is rejected, being NLS or OLS. Applied to gravity equations, the test is then non-conclusive.

5.6. Conclusion

I have shown here that the reverse of Santos Silva and Tenreyro (2006)'s proposition is also true: OLS estimates of log-linearised models are biased in the case where the true data generating process is in levels but non-linear estimates in levels are also biased in the case where the true data generating process is log-linear. This evidence is supported by Monte Carlo simulations. Since in applied cases, there is no *a priori* econometric reason to specify a model in levels rather than in logs, I try to implement some specification test to discriminate between the two. In order to do so, I use the specification test for non-nested alternative hypotheses of MacKinnon et al. (1983). Sadly, I show by Monte Carlo simulations that this test can only be used if the residuals of alternative models are both normally and identically distributed. Indeed, the test is mainly non-conclusive or misleading in other cases. Finally, I try to implement this test to discriminate between log-linear OLS and non-linear least squares estimates in levels of a gravity equation but it is non-conclusive. Future works in this direction should look for a way to satisfactorily discriminate between a non-linear estimator in levels and the OLS estimator of its log-linearised counterpart.

APPENDIX

5.A. Augmented speficification

$$X_1 \sim \mathcal{N}(0,1) \tag{5.19}$$

$$X_2 \sim \mathcal{B}(0.4) \tag{5.20}$$

$$\beta_0 = 0 \qquad \text{and} \qquad \beta_1 = \beta_2 = 1 \tag{5.21}$$

5.A.1. Specification in levels

The first set of simulations is based on the additional following assumptions.

$$E[y_i|x] = \mu(x_i\beta) = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \tag{5.22}$$

$$y_i = \mu(x_i\beta)\eta_i \tag{5.23}$$

where η_i is defined as in section 5.3.1.

Table 5.6: Monte Carlo simulations, 10,000 replicas

	β_0		β_1		β_2	
Estimator	Bias	S.E.	Bias	S.E.	Bias	S.E.
	Case 1: $V[y_i\|x] = 1 \longrightarrow$ NLS optimal					
PPML	-0.00049	0.03525	0.00026	0.01581	-0.00005	0.02713
GPML	-0.00815	0.07380	0.01344	0.06824	0.00739	0.08321
NLS	0.00004	0.02104	0.00004	0.00763	-0.00027	0.01728
OLS	-0.53350	0.04335	0.39030	0.03940	0.35568	0.05411
	Case 2: $V[y_i\|x] = \mu(x_i\beta) \longrightarrow$ PPML optimal					
PPML	-0.00115	0.03747	0.00001	0.01947	0.00082	0.04004
GPML	-0.00347	0.05075	0.00439	0.04322	0.00296	0.06282
NLS	-0.00320	0.06876	0.00057	0.03280	0.00131	0.05864
OLS	-0.40383	0.03682	0.21122	0.02959	0.20077	0.04908
	Case 3: $V[y_i\|x] = \mu(x_i\beta)^2 \longrightarrow$ GPML optimal					
PPML	-0.00050	0.08055	-0.00363	0.07057	-0.00266	0.10201
GPML	-0.00145	0.04054	0.00040	0.03122	-0.00070	0.06515
NLS	-0.46447	7.77296	0.13008	2.29375	0.07946	2.05994
OLS	-0.34689	0.03371	0.00036	0.02618	0.00020	0.05392

S.E. = Standard Error. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

5.A.2. Specifications in logs

The second set of simulations is based on the additional following assumptions.

$$E[log\ y_i|x_i] = x_i\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{5.24}$$

$$log\ y_i = x_i\beta + \epsilon_i \tag{5.25}$$

where ϵ_i is defined as in section 5.3.2.

$$\text{PPML, GPML, NLS: } y_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i) \tag{5.26}$$

Table 5.7: Monte Carlo simulations, 10,000 replicas

	β_0		β_1		β_2	
Estimator	Bias	S.E.	Bias	S.E.	Bias	S.E.
			Case 4: $\sigma_i^2 = 1$ (homoskedastic)			
PPML	0.49407	0.10904	-0.00368	0.09426	-0.00047	0.13149
GPML	0.49784	0.05354	-0.00143	0.04062	-0.00142	0.08305
NLS	-0.86143	53.25341	0.40018	15.99288	0.23328	10.72797
OLS	0.00006	0.04081	-0.00110	0.03145	-0.00096	0.06463
			Case 5: $\sigma_i^2 = \mu(x_i\beta)$ (heteroskedastic)			
PPML	-5.45072	50.41271	3.27851	15.67578	3.63585	11.01570
GPML	1.10774	0.40622	1.15335	0.33883	1.52085	0.96679
NLS	-31.70139	215.35942	11.47522	70.17809	5.65049	34.64162
OLS	-0.00082	0.05228	0.00006	0.07489	0.00103	0.11807

S.E. = Standard Error. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

5.A.3. Specification testing

Table 5.8: Test levels VS logs, 10,000 replicas

	1%		5%		10%	
Estimator	% Ok	% N.C.	% Ok	% N.C.	% Ok	% N.C.
	Case 1: $V[y_i\|x] = 1 \longrightarrow$ NLS optimal					
PPML	0.976	0.024	0.970	0.030	0.929	0.071
GPML	0.956	0.043	0.947	0.053	0.917	0.083
NLS	0.976	0.024	0.970	0.030	0.930	0.070
	Case 2: $V[y_i\|x] = \mu(x_i\beta) \longrightarrow$ PPML optimal					
PPML	0.087	0.896	0.379	0.569	0.562	0.378
GPML	0.052	0.926	0.337	0.604	0.547	0.383
NLS	0.091	0.895	0.385	0.570	0.570	0.378
	Case 3: $V[y_i\|x] = \mu(x_i\beta)^2 \longrightarrow$ GPML optimal					
PPML	0.005	0.640	0.017	0.515	0.029	0.454
GPML	0.003	0.659	0.015	0.543	0.029	0.488
NLS	0.005	0.732	0.015	0.627	0.030	0.587
	Case 4: $\sigma_i^2 = 1$ (homoskedastic)					
PPML	0.319	0.675	0.444	0.534	0.497	0.468
GPML	0.301	0.694	0.416	0.565	0.463	0.503
NLS	0.224	0.770	0.327	0.651	0.375	0.588
	Case 5: $\sigma_i^2 = \mu(x_i\beta)$ (heteroskedastic)					
PPML	0.164	0.459	0.207	0.468	0.216	0.508
GPML	0.205	0.536	0.220	0.542	0.214	0.568
NLS	0.048	0.471	0.127	0.421	0.171	0.446

PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares. N.C. = Non Conclusive.

5.B. The traditional gravity equation

Table 5.9: The traditional gravity equation

	OLS	NLS > 0	NLS	PPML > 0	PPML	GPML > 0	GPML
Log exporter's GDP	0.938^a	0.738^a	0.738^a	0.721^a	0.732^a	0.669^a	0.978^a
	(0.012)	(0.038)	(0.038)	(0.027)	(0.027)	(0.022)	(0.034)
Log importer's GDP	0.798^a	0.862^a	0.862^a	0.732^a	0.741^a	0.654^a	0.859^a
	(0.012)	(0.050)	(0.050)	(0.028)	(0.027)	(0.018)	(0.026)
Log exporter's GDP per capita	0.207^a	0.395^b	0.396^b	0.154^a	0.157^a	0.222^a	0.297^a
	(0.017)	(0.172)	(0.171)	(0.053)	(0.053)	(0.024)	(0.035)
Log importer's GDP per capita	0.106^a	-0.033	-0.033	0.133^a	0.135^a	0.158^a	0.254^a
	(0.018)	(0.056)	(0.056)	(0.044)	(0.045)	(0.024)	(0.041)
Log distance	-1.166^a	-0.923^a	-0.924^a	-0.776^a	-0.784^a	-0.954^a	-1.477^a
	(0.034)	(0.090)	(0.090)	(0.055)	(0.055)	(0.047)	(0.080)
Contiguity dummy	0.314^b	-0.081	-0.081	0.202^c	0.193^c	0.401^b	0.204
	(0.127)	(0.110)	(0.110)	(0.105)	(0.104)	(0.189)	(0.376)
Common-language dummy	0.678^a	0.689^a	0.689^a	0.751^a	0.746^a	0.646^a	0.692^a
	(0.067)	(0.089)	(0.089)	(0.134)	(0.135)	(0.091)	(0.129)
Colonial-tie dummy	0.397^a	0.036	0.036	0.020	0.025	0.185^b	0.390^a
	(0.070)	(0.135)	(0.135)	(0.150)	(0.150)	(0.093)	(0.146)
Landlocked-exporter dummy	-0.062	-1.367^a	-1.367^a	-0.872^a	-0.863^a	-0.393^a	-0.589^a
	(0.062)	(0.234)	(0.234)	(0.157)	(0.157)	(0.087)	(0.122)
Landlocked-importer dummy	-0.665^a	-0.472^b	-0.471^b	-0.703^a	-0.696^a	-0.637^a	-0.881^a
	(0.060)	(0.193)	(0.193)	(0.141)	(0.141)	(0.089)	(0.129)
Exporter's remoteness	0.467^a	1.188^a	1.188^a	0.647^a	0.660^a	0.700^a	1.020^a
	(0.079)	(0.188)	(0.188)	(0.135)	(0.134)	(0.108)	(0.177)
Importer's remoteness	-0.205^b	1.009^a	1.010^a	0.549^a	0.562^a	0.234^c	0.113
	(0.085)	(0.146)	(0.146)	(0.120)	(0.119)	(0.128)	(0.240)
Free-trade agreement dummy	0.491^a	0.443^a	0.443^a	0.179^b	0.181^b	0.621^a	1.549^a
	(0.097)	(0.134)	(0.134)	(0.090)	(0.089)	(0.199)	(0.533)
Openness	-0.170^a	0.927^a	0.928^a	-0.139	-0.107	-0.415^a	-0.414^a
	(0.053)	(0.242)	(0.241)	(0.133)	(0.131)	(0.075)	(0.104)
Observations	9613	9613	18360	9613	18360	9613	18360

Standard error between brackets. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

Table 5.10: Test on the traditional gravity equation

	without zeros		with zeros	
	(1)	(2)	(3)	(4)
	H_0: NLS	H_0: OLS	H_0: NLS	H_0: OLS
	H_1: OLS	H_1: NLS	H_1: OLS	H_1: NLS
Alpha	34426.109[a]	0.000[a]	34578.361[a]	0.000[a]
	(8702.221)	(0.000)	(8707.346)	(0.000)
Derivatives	Yes	Yes	Yes	Yes
Observations	9613	9613	9613	9613

Standard error between brackets. [a] Significant at the 1% level. PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, GPML = gamma PML, NLS = non-linear least squares, OLS = log-linear ordinaly least squares.

Chapter 6

Limitations in the Theory and the Empirics of Gravity Equations: An Analysis Based on the Pitfall of the "Distance Puzzle"[1]

This paper carries out a thorough analysis of the distance elasticity of trade between 1948 and 2006. This elasticity sharply increased when gravity equations are estimated by log OLS, while it was broadly stable based on PPML, a new benchmark in the field. We show that such a divergence is due to the increased heterogeneity of trade flows. However,

1. This chapter is a very revised version of GREQAM working paper n° 2009-12 co-written with Hervé Boulhol. We are particulary grateful to Thierry Mayer who provided data as well as useful suggestions at an early stage. We also would like to greatly thank Joao Santos Silva, Pierre-Philippe Combes and Lionel Fontagné for useful comments, as well as participants of the GREQAM PhD Students lunch seminar, the 2009 RIEF doctoral meetings, the 2009 EEA congress, the 2009 AFSE congress and the 2009 ASSET congress.

gamma PML, which should be consistent under the assumptions that make PPML so appealing, generate estimates that are significantly different from PPML and actually closer to log OLS. This raises an important challenge for the frontier knowledge of gravity specifications.

6.1. Introduction

The gravity equation of bilateral trade flows assumes a relationship that is analogous to the Newton's law of universal gravitation. While traditional specifications were largely a-theoretic, there have been major advances in the formalisation of bilateral trade flows since the mid-nineties. In an effort to lay out microfoundations, Deardoff (1998) shows that not only the bilateral distance between two countries but also their geographical positions relative to all other countries matter for the level of trade between these two countries. Consequently, many researchers since then have added a remoteness indicator to the list of explanatory variables, approximating remoteness by the weighted average of distances from all trading partners, with trading partners' GDP as weights.

The decisive methodological contribution of Anderson and van Wincoop (2003) consists of deriving an operational gravity model in which "multilateral resistance" – a function of all bilateral trade costs – is a determinant of each bilateral trade flow. They show that the absence of the multilateral resistance terms in traditional gravity estimations leads to biased estimates of some key parameters such as the effect of a common border, as these missing terms are correlated with traditional explanatory variables. Anderson and van Wincoop conclude that the remoteness variables as commonly approximated are disconnected from theory, and that replacing the multilateral resistance terms by country fixed effect leads to consistent estimates of the gravity equation, which they specify in log form.

In turn, Santos Silva and Tenreyro (2006, SST hereafter) highlight another typical bias of gravity equations that are estimated in log form, on top of the sample selection bias that results from the implied exclusion of zero trade flows. They show that heteroskedasticity in trade levels is most likely to generate biases for the log-OLS estimator as a consequence of Jensen's inequality, because the expected value of the logarithm of trade flows depends on higher moments, including the variance, which might be correlated with explanatory variables. They suggest instead that the Poisson Pseudo-Maximum Likelihood (PPML) of the gravity equations specified in level is a preferable unbiased estimator and likely to be more efficient than other estimators such as the nonlinear least squares (NLS) and gamma PML (GPML) estimators. Following this major contribution, the PPML estimator has become the benchmark in the field.

Helpman, Melitz and Rubinstein (2008) have developed a limited-dependent-variable approach in order to deal with zero flows. They use a Heckman two-step procedure relying on an identifying restriction whereby a variable determines whether trade is zero or not, but does not explain the size of trade flows. Specifically, Helpman *et al.* (2008) use common religion as the exclusion variable. In practice, it is very difficult to find a variable explaining both the probability that two countries have a trading relationship but not the size of trade between them.[2] Moreover, this approach does not deal with the heteroskedasticity issue raised by SST. Our paper sticks to comparing the estimators discussed by SST.

2. Felbermayr and Kohler (2006) use a Tobit approach, which does not need a true identifying restriction. However, their dependent variable is $log(trade + 1)$ which makes the estimates highly dependent on the unit choice for trade flows: whether trade is measured in millions of dollars or in thousands of euros makes a material difference in the estimated parameters of interests. In this case, using a Tobit model leads to the same issue of scale-dependence.

Our research project started by focusing on the analysis of the so called "distance puzzle", *i.e.* the idea that despite globalisation the elasticity of trade to distance is generally found to have increased over time (in absolute terms). To our knowledge, this is the first paper studying the "distance puzzle" using the PPML estimator. In order to simplify the text, the current paper always discusses the evolution of this elasticity in absolute values since there is no ambiguity about its sign. Taking PPML as a benchmark, we show that not only the level but also the changes in the distance elasticity through time are severely biased with log-OLS due to the *increased* heteroskedasticity over the past decades. Hence the conclusions of SST remain valid with respect to the main shortcoming of the log-OLS estimator.

Howerver, our main contribution might be to highlight some remaining important flaws in our current understanding of gravity equations. The choice between the estimators that relies on the proportionality of the conditional variance to a power of the conditional mean (PPML, NLS, GPML) of the nonlinear model specified in levels is supposed to be driven by efficiency considerations. In particular, all these estimators are supposed to be consistent and that choice guided by testing the relationship between the residuals and the estimated flows. Contrary to these predictions, point estimates differ very significantly across these estimators. These differences have such a magnitude that, given the size of our sample, they cannot be explained by finite-sample limitations. This suggests either that the theory remains unsatisfactory or that potentially important explanatory variables are generally omitted. Another result is that the test used to help identifying the most efficient estimator is weak: its inferences depend upon which estimator is firstly selected to compute the estimated trade flows and the residuals. These results should encourage trade economists to search for ways to enhance our representation of trade flows.

The rest of the paper is organised as follows. Section 6.2 introduces the "distance puzzle" and Section 6.3 discusses the main methodological and empirical issues when estimating gravity equations to analyse the "distance puzzle". Section 6.4 presents the data and the specifications, while results are shown in Section 6.5. Section 6.6 concludes.

6.2. The "distance puzzle"

Despite globalization, the role of distance in shaping world trade across trading partners seem to have increased over time, a stylised fact framed as the "distance puzzle" by Buch, Kleinert and Toubal (2004).[3] Unfortunately, except in Coe, Subramanian and Tamirisa (2007) discussed below, the estimated change in the elasticity of trade with respect to distance has been based on gravity equations specified in logarithm of trade flows, which are subject to the biases highlighted by SST. Moreover, it is not clear in the first place why a non-decreasing elasticity would represent a puzzle. The rather vague presumption seems to be that the expansion of world trade associated with a fall in distance-related trade costs means that distance is having a lesser impact on the structure of trade. Noting that the elasticity of trade to distance is the product of the elasticity of trade to trade costs and of the elasticity of trade costs to distance, there are several reasons to be skeptical.

First, in a careful formalisation of gravity equation, Anderson and van Wincoop (2003) show that trade is actually a homogenous function of degree zero in trade costs due to the multilateral resistance terms. Therefore, even though a general decrease in tariffs spurs

3. Indeed, according to the meta-analysis carried out by Disdier and Head (2008), trade decreases with distance by at least the same amount today than thirty years ago, with an increase in the (absolute value of the) distance elasticity of trade since the late eighties.

international relative to domestic trade, a uniform decrease in transport costs might not lead to increased trade. Second, the idea that a non-decreasing total elasticity represents a puzzle seems somehow related to the "world-is-getting-flatter" hypothesis that Leamer (2007) questions, pointing out that trade remains mostly a neighbourhood phenomenon, as long-distance flows seem to have increased less than short-distance ones. Third, an overall decrease in transport costs does not necessarily imply a lower distance elasticity of trade. For example, if trade costs, τ_{ijt}, between countries i and j in year t is a function of the bilateral distance d_{ij} such as $a_t\ d_{ij}^{\gamma_t}$, distance can become irrelevant over time through either a decrease in a_t or in γ_t, but a uniform decrease in distance-related transport costs is associated with a fall in a_t with no implication for the elasticity γ_t (Buch *et al.*). Even the presumption that transport costs have declined relative to the price of the goods being transported, *i.e.* mostly manufacturing goods, is far from obvious according to recent studies that provide direct measures of costs over different routes and modes of transportation (Hummels, 2007; Golub and Tomasik, 2008). Finally, the elasticity of trade with respect to trade costs might have rather increased. Based on theory (*e.g.* Anderson and van Wincoop), this elasticity (in absolute terms) is positively related to the elasticity of substitution between varieties, and it is often believed that globalisation is associated with an increase in the degree of substitutability between varieties, thereby inducing an increase in the elasticity of trade to distance. Berthelon and Freund (2008) provide evidence that the increase in the importance of distance over time is related to the substitutability of goods. Notwithstanding our conceptual skepticism, the analysis of how the elasticity of trade to distance has evolved in past decades is interesting in its own right.

Coe et al. (2007) is the paper that went the furthest in this analysis. They argue that taking into account zero trade flows enables to resolve the puzzle: they find that the elasticity of trade to distance significantly decreased from roughly 0.5 in 1975 to 0.3 in 2000. Even though Coe *et al.* bring an important contribution highlighting the importance of a nonlinear specification of gravity equations, there is wide scope for improving the analysis of the "distance puzzle" First, SST argue that PPML is likely to be much more efficient than NLS, while Coe *et al.*'s main result is established using NLS. A robustness check is performed using a Pseudo-Maximum Likelihood, but it is not totally clear which one is used. Second, the sample is restricted to 73 countries and start in 1975 only. Third, the sum of exports and imports is used as the dependent variable, but Baldwin and Taglioni (2006) warn that exports and imports must be distinguished. Fourth, the data used for Free Trade Agreements (FTAs) are not well defined. However, because FTAs have mainly promoted regional integration, they are *de facto* inversely related to distance. Therefore, not properly controlling for FTAs might be misleading.

6.3. The empirics of gravity equations

6.3.1. Cross section: consistency, efficiency and competing estimators

SST highlight a typical bias of gravity equations that are estimated in log form, on top of the sample selection bias that results from the implied exclusion of zero trade flows. Starting from a stochastic version of the gravity equation in levels, the log-linear specification generates biases as a consequence of Jensen's inequality ($E(ln\ x) \neq ln\ E(x)$), because the expected value of the logarithm of trade flows depends on higher moments, including the variance. Formally,

$$x_{ij} = exp(Z_{ij}\beta)u_{ij} \quad , \quad E(u|Z) = 1 \tag{6.1}$$

$$Var(u|Z) \neq 0 \;\Rightarrow\; E(ln\ u|Z) \neq 0 \tag{6.2}$$

where the Z explanatory variables include importer and exporter fixed effects, (log of) bilateral distances and other control variables influencing trade costs. Since the *variance* of the residuals is likely to depend on explanatory variables such as importer and exporter characteristics (which include observed ones such as GDP), estimators using the log specification might bias the parameters of interest.[4] Thus, the magnitude of the bias depends on the structure of the variance of the residuals, and heteroskedasticity in the trade level equation could invalidate inferences made from estimates based on the log-linear specification.

This problem can be overcome by estimating the level equation (6.1) using a nonlinear estimator. SST propose the PPML estimator, assuming that the variance of x is proportional to its conditional expectancy, and argue that this estimator is likely to be more efficient than NLS. Indeed, it is unrealistic to assume, as implicit with NLS, that the variance of estimated trade flows is the same for small/remote and large/central countries. Whatever the specific choice of the (nonlinear) estimator, a level specification allows for the inclusion of zero trade flows, even though SST show *ex post*, based on the empirical analysis, that including the zero flows does not make a material difference.

4. Indeed, in that case, the conditional variance depends on Z, and the bias is not limited to the constant (see the last paragraph of this sub-section).

A natural extension consists in assuming other distributions than Poisson. This can include gamma distribution according to which the variance is proportional to the square of the conditional mean, and more generally any power of it. Some authors have also used the negative binomial distribution (*e.g.* Head, Mayer and Ries (2009)), but this is inadequate when applied to trade flows because such an estimator artificially depends on the unit chosen for the dependent variable (Bosquet and Boulhol, 2010).

In order to discriminate between the various *a priori* legitimate PML estimators and select the most efficient, Manning and Mullahy (2001) suggest that if $Var(x_{ij}|Z) = \lambda_0\ E(x_{ij}|Z)^{\lambda_1}$, the choice of the appropriate estimator can be based on an asymptotically valid estimate of λ_1 from:

$$(x_{ij} - \tilde{x_{ij}})^2 = \lambda_0\ \tilde{x_{ij}}^{\lambda_1} + \zeta_{ij} \tag{6.3}$$

where $\tilde{x_{ij}}$ is the value of $E(x_{ij}|Z)$ estimated from an initially consistent estimator, such as PPML. [5]

Finally, some properties of the estimator of gravity equations in logarithm are worth mentioning. A Taylor series that is limited to the second moment around the conditional mean gives:

$$Log\ x_{ij} \approx Log\ E(x_{ij}|Z) + \frac{x_{ij} - E(x_{ij}|Z)}{E(x_{ij}|Z)} - \frac{1}{2}\frac{[x_{ij} - E(x_{ij}|Z)]^2}{E^2(x_{ij}|Z)}$$

5. SST actually suggest testing the adequacy of a particular value of λ_1 from a Taylor expansion of (6.3), which they apply in the empirical part of their paper. Unfortunately, this procedure is subject to the same problem as for the negative binomial estimator: it artificially depends on the unit choice of trade flows, and could therefore be misleading.

And therefore [6],

$$E(Log\ x_{ij}|Z) \approx Log\ E(x_{ij}|Z) - \frac{1}{2}\frac{Var(x_{ij}|Z)}{E^2(x_{ij}|Z)} \tag{6.4}$$

$$Var(Log\ x_{ij}|Z) \approx Var(x_{ij}|Z)/E^2(x_{ij}|Z) \tag{6.5}$$

On top of the possible selection bias due to the elimination of zero trade flows, these equations summarise two issues raised by the "log of gravity". Equation (6.4) highlights the bias emphasised by SST. Beyond that bias, equation (6.5) shows that assuming errors of the log specification are i.i.d., as implicit when using OLS, is consistent with the conditional variance of the flow being proportional to the square of the conditional mean, *i.e.* with the gamma distribution. Therefore, abstracting from the sample selection bias, if the true distribution were gamma, estimating the log level equation using OLS would only bias the intercept (see eq. 6.4), and not the other parameters of interest such as the distance coefficient. In that sense, the magnitude of the biases (except the constant) of the gravity equation that is estimated using the log-linear specification depends on how far the distribution of trade flows is from the gamma distribution. If, however, the true distribution is Poisson, (6.4) and (6.5) become respectively, α being a constant:

$$E(Log\ x_{ij}|Z) \approx Log\ E(x_{ij}|Z) - \alpha/E(x_{ij}|Z)$$

$$Var(Log\ x_{ij}|Z) \approx 2\alpha/E(x_{ij}|Z)$$

In that case, the bias would be very severe for small flows. Moreover, OLS estimates of the log specification would ignore that the variance of the log is very large for small flows;

6. The computation of the variance uses the Taylor series at the first order (Delta method).

in other words, it would give far too much weight to small flows.[7]

6.3.2. Panel estimates and the "distance puzzle"

As argued by Baier and Bergstrand (2007), cross section estimates of gravity equation might be biased due to the endogeneity of free trade agreements (FTAs). Properly controlling for the influence of FTAs might be especially important for the estimate of the impact of distance. Indeed, FTAs cover an increasing share of world trade (slightly more than 30% of the total amount of flows in 2005, compared with 23% in 1985 and 15% in 1965) and are often agreements between neighbouring countries, hence an obvious correlation with distances. However, finding an instrument for FTAs that does not influence trade by any other channel is extremely difficult. In this context, Baier and Bergstrand argue that country pair idiosyncrasies should be accounted for via so-called "dyadic fixed effects" to eliminate the bias due to the endogenity of FTAs; more generally panel specifications of gravity equations make it possible to control for a battery of fixed effects. Multilateral resistance terms can be controlled for by the inclusion of origin and destination country dummies for each year.

6.4. Data and econometric specification

6.4.1. Data

Trade flow data come from the IMF Direction of Trade Statistics (DOTS) database. This database provides trade flows for a long period, starting in 1948, which is well adapted

7. In contrast, the NLS estimator of the trade level specification, although consistent, is inefficient in that case because it does not give enough weights to small flows.

to study the distance puzzle, and for 205 trade partners. Moreover, DOTS includes zeros and differentiates them from missing values. The share of zero trade flows decreased from 80% in 1948 to 29% in 2006. The sample of striclty positive trade flows, used to compare the different estimators with log-OLS, has about 3,700 flows in 1948 and 22,000 in 2006.

The geographical variables (distance between countries, common border, common language and colonial linkage dummies) are taken from the CEPII database.[8] The FTA data is broadly the same as the one used in Baier and Bergstrand (2007). Specifically, the database used by these authors has been corrected and improved by Fontagné and Zignago (2007) in their re-estimation of the impact of FTAs.[9] The proportion of the value of world trade covered by FTAs goes from 7% in 1958 to 31% in 2006.

6.4.2. Specification

Following the discussion in Section 6.2, the gravity equation is estimated in levels including importer and exporter fixed effects. Formally in the cross section analysis, the following equation is estimated for each year:

$$x_{ij} = exp(\alpha_0 + \gamma \; ln \; d_{ij} + \alpha_1 \; B_{ij} + \alpha_2 \; L_{ij} + \alpha_3 \; C_{ij} + \alpha_4 \; FTA_{ij} + FX_i + FM_j)u_{ij} \quad (6.6)$$

with $Var(x_{ij}|Z) = \lambda_0 \; E(x_{ij}|Z)^{\lambda_1}$. x_{ij} is the nominal US$ value of export from i to j, FX_i and FM_j are the fixed effects for exporting and importing countries, respectively. B_{ij}, L_{ij} and C_{ij} are the traditional control covariates: common border, common official language

8. http://www.cepii.fr/anglaisgraph/bdd/distances.htm, Centre d'Etudes Prospectives et d'Informations Internationales.

9. Compared with Fontagné and Zignago, FTA data has been updated beyond 2000. In total, 47 FTAs are covered. The first FTA in the database is the European Economic Community. Its treaty was signed on March 25th, 1957, so it begins in 1958 in the database.

and colonial linkage dummies, respectively and u_{ij} are the error terms of the nonlinear estimates. In order to test the efficiency of the NLS, Poisson or gamma assumption, the most efficient power of the conditional mean is computed according to equation (6.3). The log version is also estimated using OLS.

In order to separate the various factors influencing the analysis of the distance puzzle, the gravity equations are first estimated without controlling for FTAs (FTA_{ij}). Because these first results might be subject to omitted variable biases, the analysis focuses in a second step on the impact of controlling for FTAs. Following the discussion in subsection 6.3.2, a panel specification including both time-varying importer and exporter fixed effects as well as dyadic fixed effects is estimated:

$$x_{ijt} = exp(\gamma_t \; ln \; d_{ij} + \alpha_{1t} \; B_{ij} + \alpha_{2t} \; L_{ij} + \alpha_{3t} \; C_{ij} + \alpha_{4t} \; FTA_{ijt} + FX_{it} + FM_{jt} + Dyadic_{ij}) u_{ijt} \tag{6.7}$$

With dyadic fixed effects, the parameters related to variables that are time-invariant such as geographic characteristics, including the bilateral distance, are wiped out; only the evolution through time of these parameters can be estimated. [10] It is clear that introducing the $i \times j$, $i \times t$, $j \times t$ fixed effects in a nonlinear specification generates computational difficulties. As a result, the estimation of equation (6.7) is carried out using 5-year averaging. [11] Also, in order to drastically reduce the number of dyadic fixed effects, the largest possible balanced panel was constructed; it consits of the same 2,550 pairs of countries between 1952 and 2006 covering 90 countries and 78% of world trade on average.

10. Unlike Baier and Bergstrand, elasticities are here allowed to vary through time.

11. To be consistent with the gravity specification in levels, the geometric mean of trade flows is used as the dependant variable. By comparison, Baier and Bergstrand use a specification in logs with elasticities with respect to distance, border, colonial link, etc., that are constant through time, and reduce the number of fixed effects by keeping only one out of five years.

6.5. Empirical results

This section presents first the cross-section results obtained without controlling for free trade agreements. The focus is on the elasticity of trade with respect to bilateral distance. Sub-section 6.5.1 compares PPML with log-OLS while sub-section 6.5.2 highlights the differences across the non-linear estimators. Finally, sub-section 6.5.3 provides the results obtained when controlling for FTAs in cross-section, and then using a panel specification.

6.5.1. PPML vs log-OLS, no FTAs

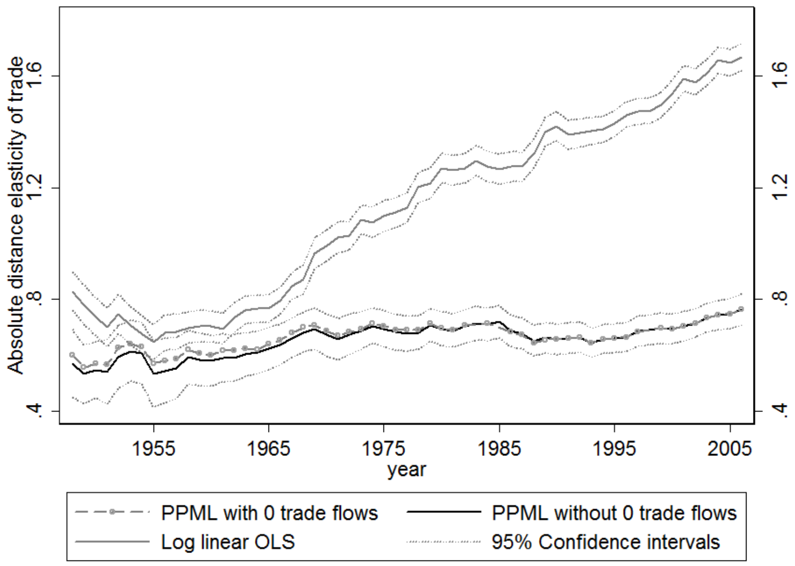
Table 6.1: Gravity equations estimated with PPML, cross sections

Year	*1955*	*1965*	*1975*	*1985*	*1995*	*2005*
Log distance	-0.53***	-0.62***	-0.70***	-0.72***	-0.66***	-0.75***
	(0.06)	*(0.04)*	*(0.03)*	*(0.03)*	*(0.03)*	*(0.03)*
Contiguity	0.57***	0.32***	0.35***	0.33***	0.65***	0.43***
	(0.19)	*(0.11)*	*(0.11)*	*(0.09)*	*(0.10)*	*(0.09)*
Language	0.31**	0.40***	0.29***	0.32***	0.18**	0.20**
	(0.12)	*(0.09)*	*(0.08)*	*(0.07)*	*(0.08)*	*(0.08)*
Colony	1.05***	0.85***	0.53***	0.14	0.11	0.16
	(0.15)	*(0.12)*	*(0.12)*	*(0.11)*	*(0.10)*	*(0.12)*
Country FE	yes	yes	yes	yes	yes	yes
Observations	4558	7449	11649	13063	19973	22201

A gravity equation is estimated for each year, where the dependant variable is the level of bilateral trade flows. Standard errors are in parentheses; ***, ** and * are significance levels at the 1%, 5% and 10% thresholds, respectively.

The gravity equation as specified by equation (6.6) is first estimated for each year by PPML. Table 6.1 presents the results for six specific years between 1955 and 2005. Based on this estimator, the elasticity of trade with respect to distance increased between 1955 and 1970 and was broadly stable after that within a 0.65-0.75 range. This range is tight compared to those found in the literature based on log specifications. The estimated robust standard error steadily declined from 0.040 to 0.025 while the size of the sample was multiplied fivefold.

Figure 6.1: Evolution of the distance elasticity of trade, PPML vs log-linear estimates, cross sections

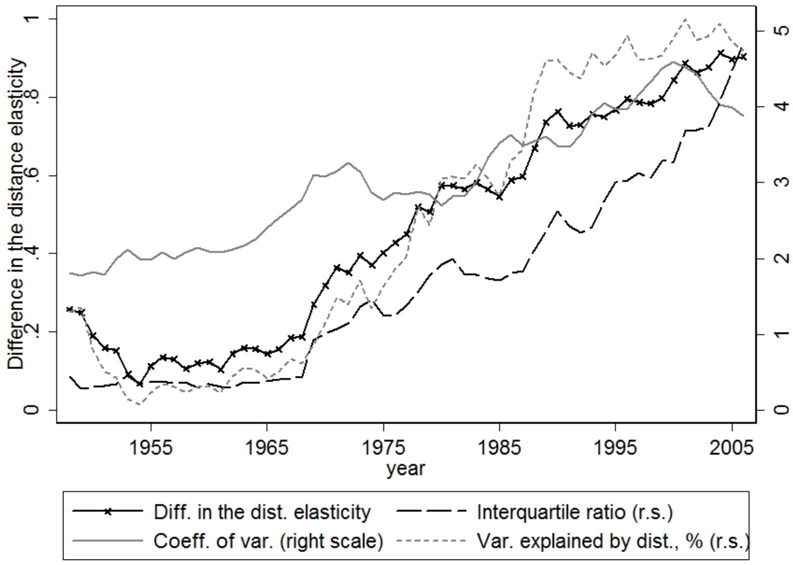


A gravity equation is estimated for each year both in levels with the PPML estimator and in log with OLS.

Figure 6.1 compares the evolution of the distance elasticity estimated using either OLS in logs and PPML in levels (including or excluding zeros) along with confidence intervals. According to the log-linear specification, the elasticity steadily increased from about 0.70 to 1.60, which characterises the distance puzzle. As a result, the difference between "PPML" and "log-linear" elasticities dramatically increased over time. In addition, the PPML distance parameter is not sensitive to whether the zero trade flows are included or not (confidence intervals are also similar), a result also found by SST, and Coe *et al.* using NLS.

This increasing difference seems to be due to the *growing* heterogeneity of flows and induced heteroskedasticity, consistent with the idea introduced by SST. The intuition behind such a link is illustrated as follows. Two measures of dispersion and one of heteroskedasticity were computed. The measures of dispersion are the inter-quartile ratio (ratio of 3rd to 1st quartile) of trade flows, and the coefficient of variation (standard deviation divided by mean). They are computed on the sample on which the log-linear specification is based, *i.e.* without zeros (again, the inclusion of zero flows has minor effects on the PPML estimates). The measure of heteroskedasticity related to the bias of the log-linear estimator is the share of the variance of $log\ \hat{u}$ explained by (the log of) distance, where $\hat{u}$ is the PPML estimated multiplicative residual. Indeed, according to Jensen's inequality, the bias of log OLS is due to the dependence of $log\ u$ on Z.[12]

Figure 6.2: Illustration of the heteroskedasticity issue



The difference in the distance elasticity is the gap between Poisson Pseudo-Maximum Likelihood and OLS log estimates of the distance elasticity of trade. The interquartile ratio if the ratio of the third over the first quartile of trade flows. The coefficient of variation is the standard deviation of trade flows divided by the mean. The variance of $log\ \hat{u}$ explained by distance is computed as the difference between the adjusted-R^2 of the regression of $log\ \hat{u}$ on the explicative variables and the adjusted-R^2 of this same regression omitting (the log of) distance as explanatory variable. To fit the right scale, the interquartile ratio and the coefficient of variation have been divided by 70 and 2.5, respectively.

12. The log of the residual is directly related to the bias as equations (6.1), (6.4) and (6.5) imply that $E(log\ u|Z) \approx -\frac{1}{2}Var(u|Z)$.

Figure 6.2 represents these three indicators in addition to the difference in the distance elasticity between PPML and log-linear OLS. Over the period, the q3 / q1 ratio has increased by a factor of 12. This is due to the tremendous increase in small non-zero flows, as q1 decreased from \$ 2 M in the 1950s to \$ 0.1 M since the mid-1990s (numbers are deflated by the US GDP deflator using 2000 as the base year). Within the same period, the average flow increased from \$ 100 M to \$ 400 M, and the standard deviation increased even faster as the coefficient of variation rose from 4 to 10. Since the small non-zero flows carry a disproportionate weight in log, the increase in its share is likely to contribute heavily to the widening of the gap between the PPML and the log-linear elasticities. Visually, the change in the difference between log OLS and PPML elasticities is closely related to that of the contribution of distance to the variance of the residuals.

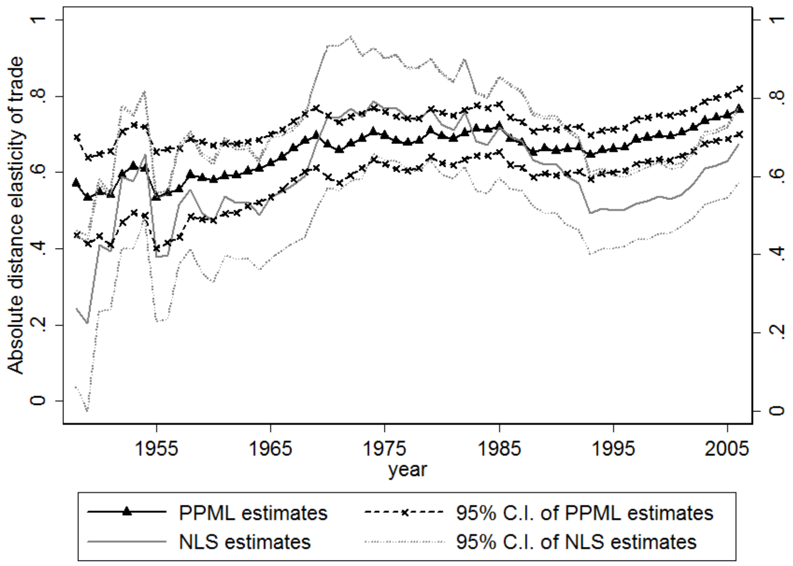
6.5.2. Differences across nonlinear estimators in level, no FTAs

This part investigates whether the assessment of the distance puzzle is sensitive to the choice of the nonlinear estimator among the class that verifies $Var(x_{ij}|Z) = \lambda_0 E(x_{ij}|Z)^{\lambda_1}$, all of them being consistent under (6.1). This includes the NLS ($\lambda_1 = 0$), the PPML ($\lambda_1 = 1$) and the GPML ($\lambda_1 = 2$).[13]

Figure 6.3 compares the NLS and PPML estimates of the distance elasticity. Levels are broadly in line except at the end of the period where the NLS elasticity is lower. Variations are also greater with NLS, and, on average, the NLS estimated standard error is twice as large as the PPML one; this might indicate that PPML is more efficient. The main difference in the evolution of the point estimates is that the elasticity fell between

13. For $\lambda_1 = 0$, NLS or maximum likelihood leads to almost identical estimates.

Figure 6.3: Evolution of the distance elasticity of trade, cross sections: PPML vs NLS estimates



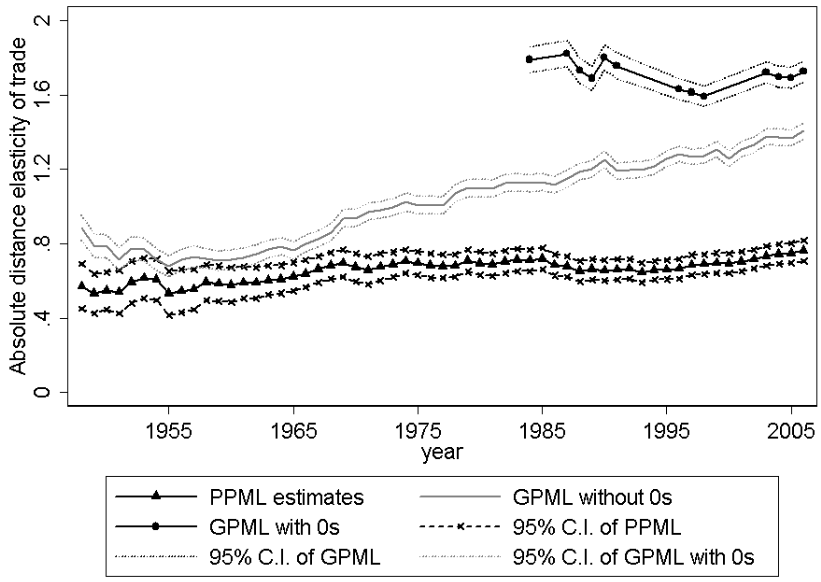
C.I. = Confidence Interval. A gravity equation is estimated in levels for each year both with Poisson Pseudo-Maximum Likelihood and Nonlinear Least Squares.

1975 and 1995 with NLS, while it was broadly stable with PPML.

The most important result is the significant difference between the elasticity estimated by PPML (or NLS) and GPML, as strikingly illustrated by Figure 6.4.[14] The trend in the GPML elasticity, without zero flows, is similar to the log-linear OLS one. The significant difference between PPML and GPML is problematic because even though GPML (as well as log-linear OLS) gives a high weight to small flows, which might be a source of poor efficiency (and bias for log OLS), both GPML and PPML should be consistent under (6.1): this is the main puzzle of the whole approach consisting in estimating gravity equations by the class of PML estimators relying on the proportionality of the conditional variance to a power of the conditional mean. Given the size of the sample, such a puzzle cannot be explained

14. The standard errors of both estimators are around 0.03, which makes them significantly different.

Figure 6.4: Evolution of the distance elasticity of trade, cross sections: PPML vs GPML estimates



C.I. = Confidence Interval. A gravity equation is estimated in levels for each year both with Poisson and Gamma Pseudo-Maximum Likelihood (with and without zero trade flows).

by finite sample limitations. Also, when zero flows are included, the GPML estimator produces estimates even further away from PPML (on top of encountering convergence problems for many years).[15]

The test presented in Section 6.2 consisting in estimating equation (6.3) to select the most efficient estimator actually proves to be weak. Indeed, its inferences depend on the choice of the estimator used to calculate $\tilde{x_{ij}}$. When PPML is used, the average estimate over the period is 1.03 with an average estimated standard error of 0.26 (Table 6.2). PPML is never rejected as optimal in that case, whereas NLS and GPML always are at the 95% confidence level. When NLS is used to calculate $\tilde{x_{ij}}$, λ_1 is estimated at 0.64 on average (average $s.e. = 0.09$), which clearly discriminates against NLS. Unfortunately, when GPML

15. The sensitivity of GPML to the inclusion / exclusion of zeros is not a problem as such since small flows are given more weight with GPML and excluding them might lead to selection bias.

Table 6.2: λ_1 is estimated from equation (6.3) using the different estimators as starting points for $\tilde{x_{ij}}$

Year	*1955*	*1965*	*1975*	*1985*	*1995*	*2005*
using PPML ($\lambda_1 = 1$)	1.33	1.06	0.80	1.07	0.96	1.06
	(0.29)	*(0.30)*	*(0.14)*	*(0.35)*	*(0.22)*	*(0.28)*
using NLS ($\lambda_1 = 0$)	0.66	0.65	0.64	0.55	0.63	0.63
	(0.12)	*(0.10)*	*(0.08)*	*(0.06)*	*(0.07)*	*(0.07)*
using GPML ($\lambda_1 = 2$)	2.33	2.92	1.93	2.58	2.16	2.10
	(0.28)	*(0.43)*	*(0.31)*	*(0.22)*	*(0.06)*	*(0.05)*
using PPML with 0s	1.36	1.12	0.80	1.08	0.96	1.06
	(0.37)	*(0.35)*	*(0.14)*	*(0.35)*	*(0.22)*	*(0.28)*
using NLS with 0s	0.66	0.65	0.64	0.56	0.63	0.63
	(0.11)	*(0.09)*	*(0.08)*	*(0.06)*	*(0.07)*	*(0.07)*
using GPML with 0s	.	.	.	2.01	2.00	2.01
	(.)	*(.)*	*(.)*	*(0.004)*	*(0.002)*	*(0.004)*

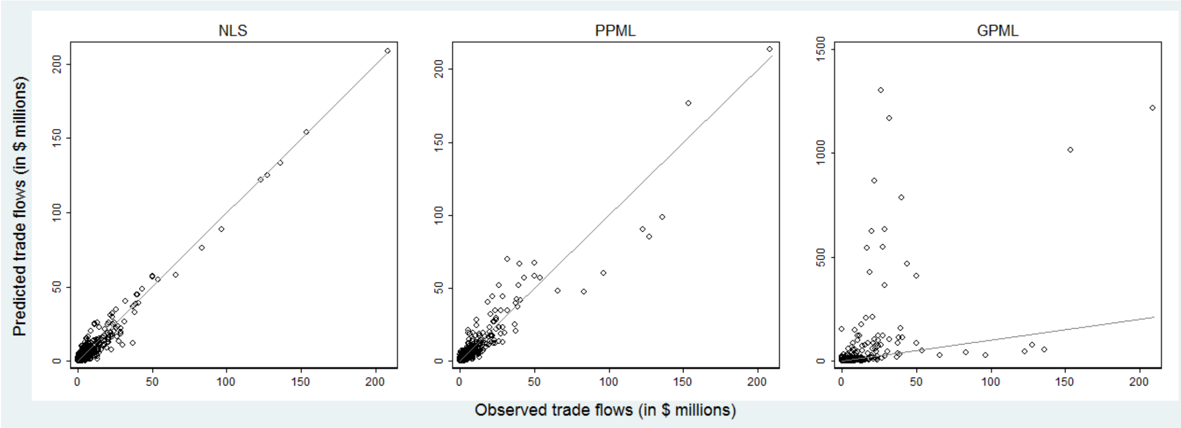
Standard errors are in parentheses. Equation (6.3) is estimated for each year. Taking the year 2005 as an example, λ_1 is estimated at 1.06 using PPML for $\tilde{x_{ij}}$, at 0.63 using NLS and at 2.10 using GPML. For GPML with 0s, years 1985 and 1995 have been respectively replaced by 1984 and 1996 because of convergence issues.

is used for $\tilde{x_{ij}}$, λ_1 is estimated at 2.31 on average (average *s.e.* = 0.22), indicating that GPML should be preferred to both PPML and NLS.

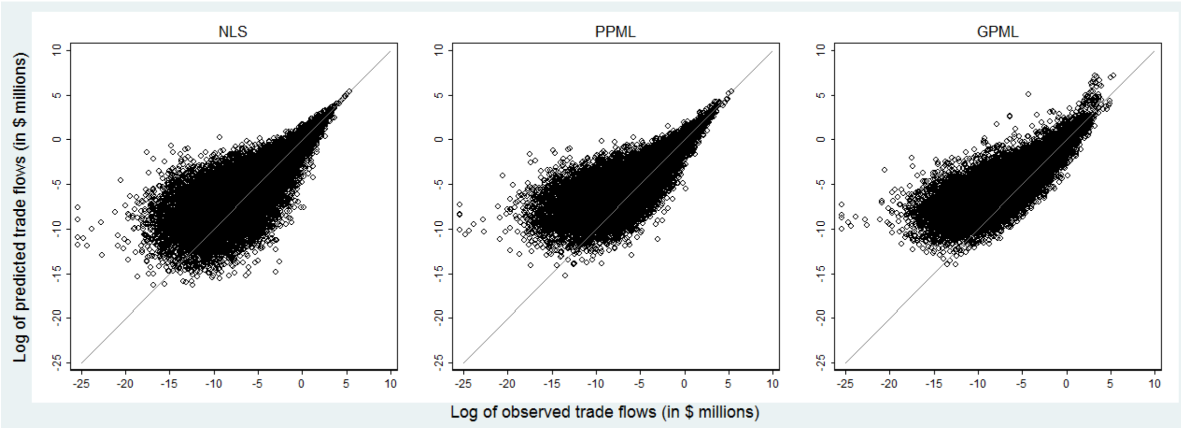
Beyond the econometric puzzle highlighted above, GPML might be a questionable estimator along two lines of argument. The first is based on a judgement call that is illustrated by taking two trade flows of \$ 10,000 and \$ 1 Billion. While NLS gives the same importance to making an estimated error of \$ 10,000 on each of them, GPML gives the same importance to making an error of say 10% on each flow, *i.e.* of \$ 1,000 and \$ 100 Million, respectively. PPML seems to be a good compromise between these two extremes. Second, Figure 6.5 presents the scatter plots of the observed flows (x-axis) and estimated ones (y-axis) using NLS, PPML and GPML. The fit for GPML is very bad. In particular, the four flows on the North-West GPML quadrant, which correspond to the bilateral flows between Belgium and Germany and between the Netherlands and Germany, are estimated by GPML to be between 35 and 50 times their actual values, and 4 to 6

Figure 6.5: Observed and predicted trade flows using different estimators, 2000

Mind the y-axis scale for GPML!



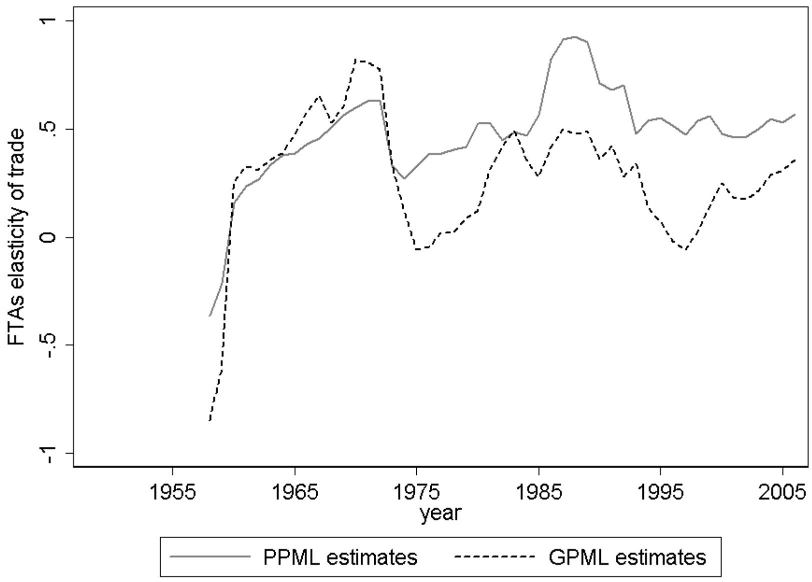
Fit in logs



The straight line is the 45ř line.

times bigger than the largest actual bilateral flow (exports from Canada to the USA). The sum of squares of the residual is equal to 4.3% of the total variance with NLS, 9.8% with PPML and 4,200% with GPML. Although NLS does better by definition on this indicator, the poor performance of GPML is extreme. We take all these as evidence against the poor performance of GPML.

Figure 6.6: Impact of FTAs in cross section analysis: Free Trade Agreements elasticity of trade, cross sections



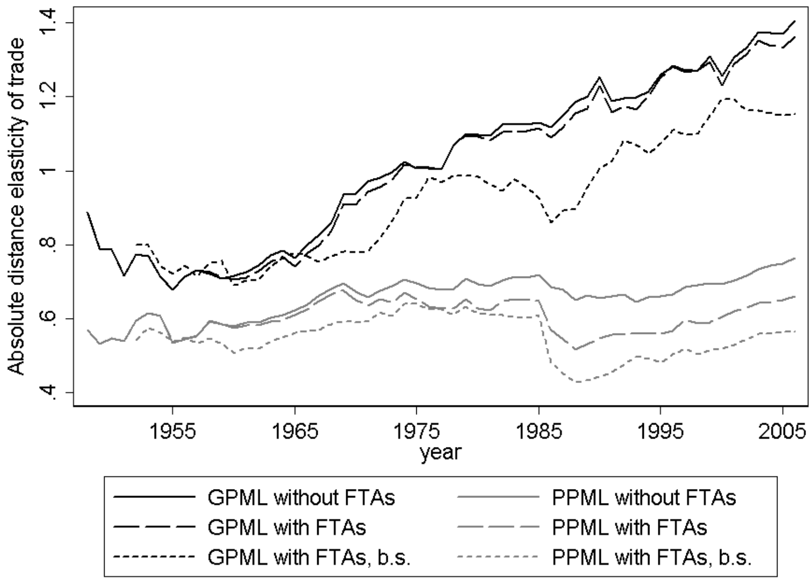
A gravity equation is estimated for each year in levels with the Poisson and gamma Pseudo-Maximum Likelihood estimators.

6.5.3. Results with FTAs, and panel specification

Accounting for FTAs has a small impact on the assessment of the "distance puzzle" in cross section. Figure 6.6 shows the estimated impact of FTAs on trade through time based on PPML and GPML in cross section. Sharing an FTA increases trade between two countries by 48% and 27%, on average through time, according to PPML and GPML, respectively. However, controlling for FTAs does not affect substantially the estimated value of the distance elasticity, even though, as expected because FTAs are negatively correlated with distance, taking into account the effect of FTAs reduces the estimated distance coefficient; the difference is never greater than 0.13 (Figure 6.7).

As argued in sub-sections 6.3.2 and 6.4.2, properly controlling for FTAs calls for introducing dyadic fixed effects and panel estimation in the $ij \times t$ dimension, which in turn

Figure 6.7: Impact of FTAs in cross section analysis: Distance elasticity of trade, cross sections

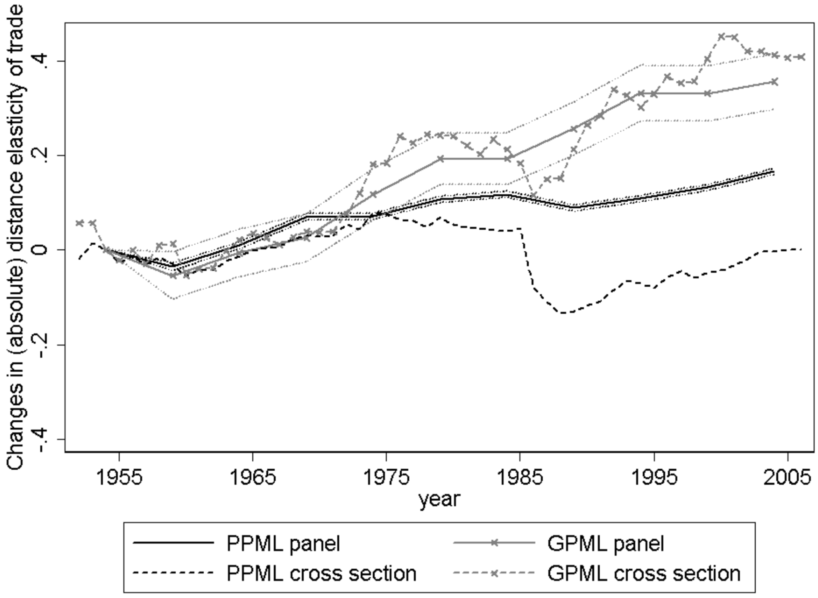


B.s. = balanced sample (see subsection 6.4.2). A gravity equation is estimated for each year in levels with the Poisson and gamma Pseudo-Maximum Likelihood estimators.

requires period averaging and working on a smaller, balanced panel. The cross section results (hence without dyadic effects) reached so far are not too sensitive to whether the analysis is carried out on the full or smaller sample, although the GPML distance elasticity is lower with the balanced panel (Figure 6.8).

The introduction of dyadic fixed effects wipes out the estimated value of the parameters related to time-invariant variables, such as the bilateral distance, but not the changes through time of such parameters. Figure 6.8 displays the panel estimates and compares them with those in Figure 6.7 (the changes are represented relative to the first period). Compared to the cross section estimates, two results are worth noting. First, the difference in the variations through time of the distance elasticity between PPML and GPML remains highly significant but is somewhat smaller. For PPML, there is a steady increase in the

Figure 6.8: Impact of FTAs in panel specification, balanced sample: Distance elasticity of trade, changes from the first period (1952-1956)



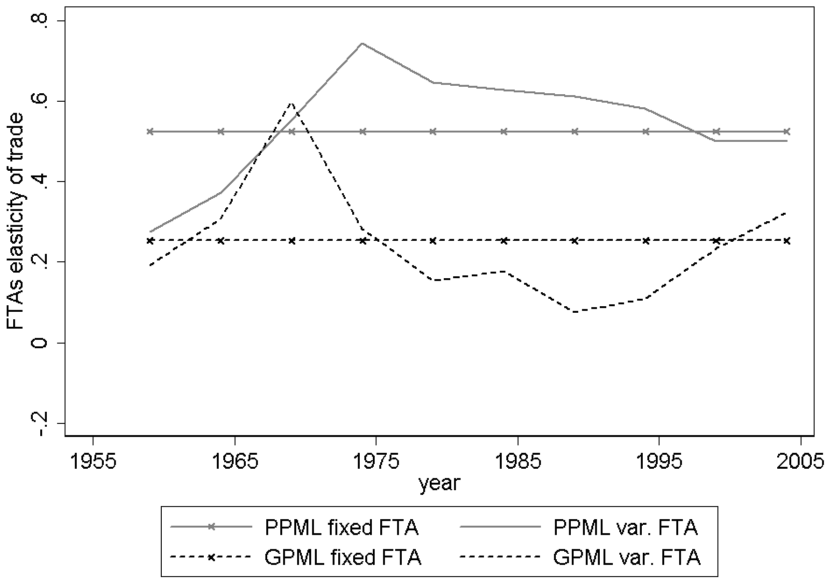
All estimates are based on the smaller balanced sample using 5-year period averaging (see subsection 6.4.2). For panel estimates the chart shows 95% confidence intervals of the changes from the first period.

elasticity since the early 1990s, albeit limited in magnitude.

Although Baier and Bergstrand constrain the elasticities to be constant over time, the panel approach allows for time-varying elasticities, as required for the analysis of the "distance puzzle". However, for comparison purposes, holding the FTA parameter constant leads with PPML (GPML) to an estimate of 0.52 (0.25) for the FTA parameter, compared with 0.49 obtained by Baier and Bergrstrand using a log specification. As shown in Figure 6.9, when the FTA parameter is allowed to be time-varying, the estimated FTA coefficient in the panel specification (eq. 6.7) varies from 0.27 to 0.74 between 1952 and 1975, and steadily decreases afterwards to 0.50. With GPML the impact of FTAs is lower on average.[16]

16. PPML and GPML lead also to differences in other parameters of interests. Sticking to PPML as the baseline, the effect of colonial linkages sharply decreases from 1.1 to about 0 over the whole period in

Figure 6.9: Impact of FTAs in panel specification, balanced sample: Free Trade Agreements elasticity of trade, panel



Fixed FTA = the parameter associated to FTAs is fixed over time; var. FTA = the parameter associated to FTAs can vary over time. All estimates are based on the smaller balanced sample using 5-year period averaging (see subsection 6.4.2).

6.6. Conclusion

We have conducted a thorough estimation of gravity equations of bilateral trade flows between 1948 and 2006 and focused on the change in the elasticity of trade with respect to the bilateral distance between trade partners. Consistent with Santos Silva and Tenreyro (2006) who estimate gravity equations for one single year, the analysis of the "distance puzzle" is likely to be severely biased when carried out by log OLS, which generates a high and increasing distance elasticity (in absolute values). In contrast, PPML – which has become the benchmark estimator – estimates suggest that the distance elasticity is lower and has been broadly stable, and we argue why a non-decreasing elasticity is not

cross-sections, while the decrease is of "only" 0.5 with PPML in panel. The effects of contiguity and of common official language are broadly stable in both cross sections and panel. Details are available upon request.

that puzzling theoretically. We also show that the increasing difference in the distance elasticity between log OLS and PPML in levels is related to the increasing heterogeneity in trade flows.

However, our results lead to a real puzzle for the estimation of gravity equations. Even though GPML might not be the most efficient estimator, it should be consistent under the assumptions that make PPML so appealing. The problem is that GPML estimates are closer to log OLS and are significantly different from PPML. The size of our sample rules out finite-sample limitations as an explanation for such a divergence between PPML and GPML.

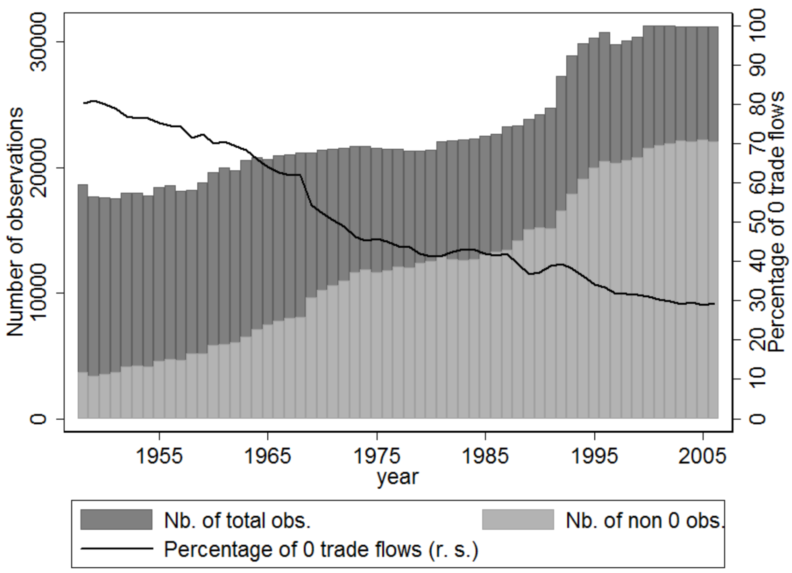
At least two possibilities arise that should draw the attention of future researchers in the field. The first one is that the theory might be inadequate, in which case it should be amended. The second one is that generally omitted explanatory variables might generate biased estimates. In any case, the important conclusion of the analysis herein is that the frontier knowledge of gravity equations might suffer from important flaws. We tried hard to solve this puzzle and reach a more constructive message, but failed in that respect.

APPENDIX

6.A. Additional descriptive statistics

Figure 6.10 plots the number of zero and non-zero trade flows through time, thereby illustrating the risk of selection bias using log-linear OLS. Indeed, despite the decreasing share of zero trade flows from 80% in 1948 to 29% in 2006, it still represents an important proportion.

Figure 6.10: Number of total trade flows and strictly positive ones in the DOTS database



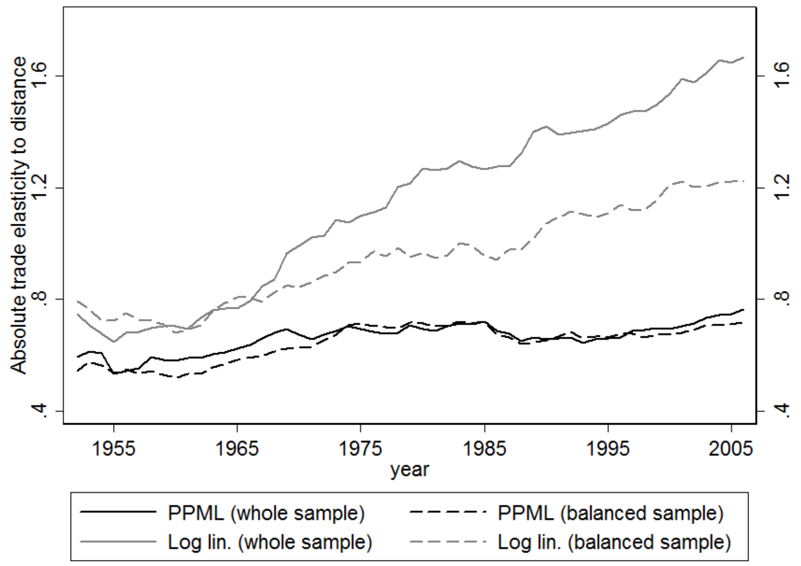
6.B. Sampling

It proved useful to work also with a balanced panel to account for the increasing number of trade flows as well as for the change in the sample over time. Hence, the largest possible balanced panel consits of the same 2550 pairs of countries between 1952 and 2006 covering

90 countries and 78% of world trade on average.

The increasing difference between OLS and PPML estimates seems to be due to the *growing* heterogeneity of flows and induced heteroskedasticity, consistent with the idea introduced by Santos Silva and Tenreyro. The intuition behind such a link is also illustrated by Figure 6.11, which replicates Figure 6.1 and adds the elasticity estimated from the smaller albeit more homogenous balanced panel. While the PPML estimate is not sensitive to the choice of the sample, the reduced heterogeneity in the balanced panel leads to a lower estimated elasticity in the log-linear specification compared with that for the whole sample, the more so for the more recent years.

Figure 6.11: Evolution of trade elasticity to geographic distance : sample analysis

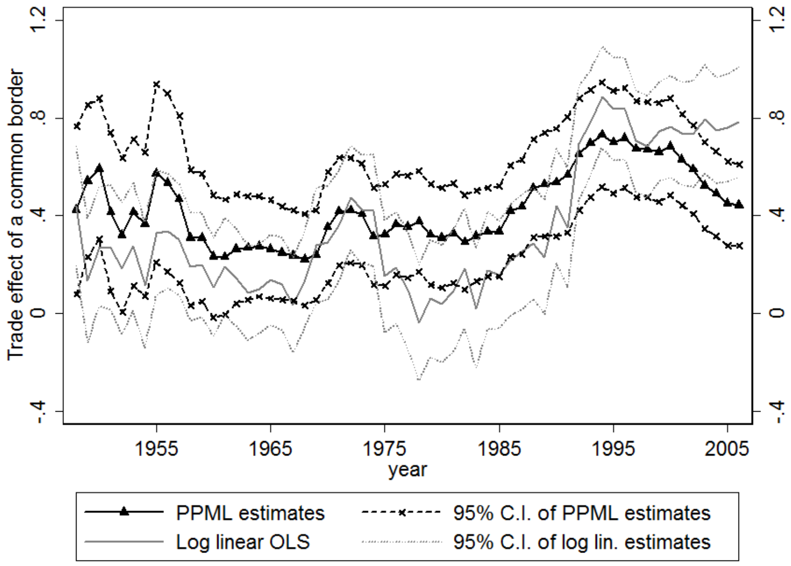


A gravity equation is estimated for each year both in levels with the PPML estimator and in log with OLS. The balanced sample contains 2550 observations.

6.C. Impact of trade determinants other than distance

In this Appendix, equation (6.6) is estimated for each year in cross section constraining the FTA parameter to 0.3. PPML and log OLS lead to similar estimates of the impact of a common border on international trade (Figure 6.12). The parameter is estimated at about 0.3 between 1948 and 1985. Then the coefficient increases for the two specifications to about 0.7 in the early nineties.

Figure 6.12: Impact of sharing a common border on trade

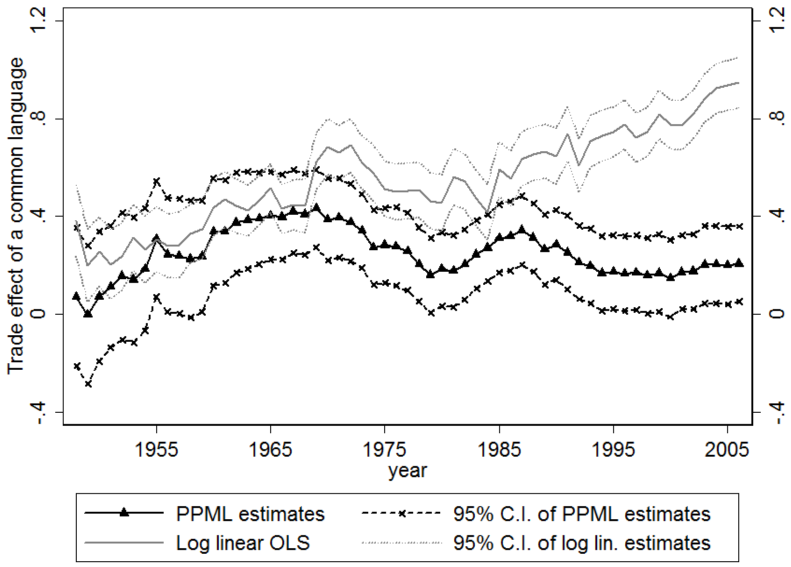


C.I. = Confidence Interval. A gravity equation is estimated for each year both in levels with the Poisson Pseudo-Maximum Likelihood estimator and in log with OLS.

The effect of a common official language is broadly stable around 0.2 when estimated by PPML (Figure 6.13). In contrast, log OLS leads to a steady increase, somehow unrealistically, of the estimated impact up to 0.95 in 2006.

The evolution of the estimated effect of having colonial linkages is illustrated in Figure 6.14. While the two methods highlight the declining importance of colonial links,

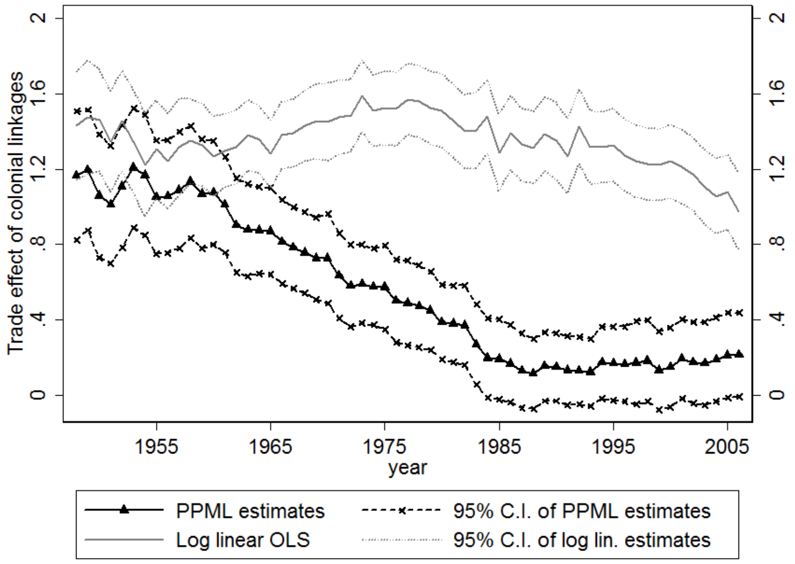
Figure 6.13: Impact of sharing a common official language on trade



C.I. = Confidence Interval. A gravity equation is estimated for each year both in levels with the Poisson Pseudo-Maximum Likelihood estimator and in log with OLS.

PPML seems to produce here also more realistic results. According to PPML, the impact of colonial linkages has been stable at about 0.2 since the mid-eighties, whereas it is still around 1.0 based on log OLS.

Figure 6.14: Impact of colonial linkages on trade



C.I. = Confidence Interval. A gravity equation is estimated for each year both in levels with the Poisson Pseudo-Maximum Likelihood estimator and in log with OLS.

Chapter 7

Applying the GLM variance assumption to overcome the scale-dependence of the Negative Binomial QGPML Estimator [1]

7.1. Introduction

Pseudo-Maximum Likelihood (PML) methods were introduced and then derived for Poisson models by Gourieroux, Monfort and Trognon (1984a,b). Following these seminal works, the Poisson PML (PPML) estimator, which assumes proportionality between the

1. This chapter is a very revised version of GREQAM working paper n° 2010-39 co-written with Hervé Boulhol. We are indebted to two anonymous referees for valuable comments and to Thierry Mayer who provided data as well as useful suggestions at an early stage. We also would like to greatly thank Joao Santos Silva, Thierry Magnac, Pierre-Philippe Combes and Lionel Fontagné, as well as participants of the GREQAM PhD Students lunch seminar, the 2009 RIEF doctoral meetings, the 2009 EEA congress, the 2009 AFSE congress and the 2009 ASSET congress.

conditional variance and the conditional expectation of the dependent variable, has often been used for count data models. However, beyond count data, Gourieroux, Monfort and Trognon (1984*b*) note that "the pseudo-maximum likelihood method with Poisson family may be applied even if the dependent variable is any real number".

Santos Silva and Tenreyro (2006) highlight the advantages of this estimator for gravity equations of bilateral trade flows specified in levels, relative to the common practice of estimating these equations in log-levels by Ordinary Least Squares. Indeed, these authors show that the log-linear specification leads to biased estimates following Jensen's inequality, due to heteroskedasticity in trade levels.[2] Moreover, they provide some evidence that the PPML estimator is likely to be more efficient than the nonlinear least squares estimator of the trade specification in levels.

As a result, a number of empirical studies of trade flows have applied the PPML estimator. As an extension, some researchers have considered other PML estimators based on non-Poisson distributions such as gamma according to which the variance is proportional to the square of the conditional mean. The Negative Binomial Quasi-Generalised PML (NB QGPML, Gourieroux *et al.*, 1984a,b) estimator is also increasingly used in trade as well as mergers and acquisitions studies, including Head et al. (2009), Burger, van Oort and Linders (2009), Briant, Combes and Lafourcade (2009), Westerlund and Wilhelmsson (2011) and Garita and van Marrewijk (2008). The NB distribution assumes that the

2. Because $E(Log\ y) \neq Log\ E(y)$, the expected value of the logarithm of trade flows depends on higher moments, including the variance. Since the conditional variance of the residuals is likely to depend on explanatory variables, estimators using the log specification might be biased. Also, as highlighted by one referee, estimating models such as gravity equations directly in levels avoids the problem of re-transforming logged outcomes back into their original scales when one wishes to make predictions of the outcome measures. Manning and Mullahy (2001) argue that in the presence of heteroskedasticity in the log-linear specification, the exponentiated log-scale prediction might provide a biased estimate of the conditional expectation.

conditional variance is equal to the conditional mean plus the product of its square and a scalar to be estimated.[3] The NB QGPML estimators that are used in these empirical analyses are two-step estimators whereby the dispersion parameter is estimated in a first step. This estimator is appealing because by encompassing both PPML and gamma PML (GPML) as special cases it is meant to increase efficiency.

This note shows that the NB QGPML estimators that have been used, although consistent among the class of exponential families, do not improve upon PPML and GPML when applied to continuous dependent variables, such as trade or M&A flows, for which the choice of the unit measure is arbitrary. This is because these NB QGPML estimates artificially cover the full range of estimates between PPML and GPML, depending only on the unit chosen. For example, in the case of trade equations, the NB QGPML estimated parameters depend on whether trade flows are measured in thousands of dollars, in billions of dollars or in millions of euros. More precisely, when flows are measured in small units (*e.g.* thousands of dollars), the NB QGPML converges towards the GPML estimator. In contrast, when flows are measured in large units (*e.g.* trillions of dollars), the NB QGPML converges towards the PPML estimator. This scale dependence of the proposed estimators has so far been unnoticed.

To solve this issue, we introduce a NB QGPML estimator, NB QGPML^{GLM}, relying on the GLM variance assumption (Wooldridge, 1999), *i.e.* assuming that the conditional variance is equal to a linear combination, to be estimated, of both the conditional mean and of its square. This new estimator is scale-invariant and likely to improve efficiency upon both Poisson and gamma PML which it truly encompasses. As this note will show, the flaw

3. Cameron and Trivedi (1986) refer to the Negbin II parameterisation of the negative binomial specification in that case.

in the usual two-step NB PML estimators comes from constraining the first-step equation, which fits the estimated conditional variance, by using the nominal variance assumption (Wooldridge, 1999), *i.e.* by constraining the coefficient on the conditional expectation to be equal to 1.

Section 7.2 provides the proof of the scale-dependence of the commonly used NB QGPML estimators and introduces our NB QGPML^{GLM} estimator overcoming this limitation. Sections 7.3 and 7.4 illustrate this proposition through an application based on Monte-Carlo simulations and on the trade gravity equation, respectively. Section 7.5 concludes.

7.2. Proof

7.2.1. Scale-dependence of usual NB QGPML estimators

The specification is $y_i = exp(X_i\ \beta)\eta_i$ where η_i is the multiplicative residual, with $E(\eta_i|X_i) = 1$.[4] The first-order conditions (f.o.c.) on β for PPML, NB PML and GPML are, respectively (Gourieroux *et al.* ; 1984a,b):

$$\text{PPML:} \quad \sum_i \big(y_i - exp(X_i\ \beta)\big)\ X_i = 0 \tag{7.1}$$

$$\text{NB PML:} \quad \sum_i \big(1 + \alpha\ exp(X_i\ \beta)\big)^{-1}\ \big(y_i - exp(X_i\ \beta)\big)\ X_i = 0 \tag{7.2}$$

$$\text{GPML:} \quad \sum_i exp(-X_i\ \beta)\ \big(y_i - exp(X_i\ \beta)\big)\ X_i = 0 \tag{7.3}$$

4. We follow here the first specification of Santos Silva and Tenreyro (2006). It is indifferent to specify the model as $y_i = exp(X_i\ \beta) + u_i$ with $E(u_i|X_i) = 0$.

Whereas the underlying assumption of PPML and GPML is that the conditional variance is proportional to the conditional expectation and to its square, respectively, the NB PML assumes that $Var(y|X) = E(y|X) + \alpha\, E^2(y|X)$, where α is a constant, generally considered to be positive. As is well known, eqs. (7.1-7.3) indicate that, when $\alpha \to 0$, NB PML $\to$ PPML, while when $\alpha \to \infty$, NB PML $\to$ GPML.

This note focuses on the impact of using $\tilde{y} = \lambda\, y$ as the dependent variable instead of y where λ is a scalar that can be either (very) small or (very) large depending on the unit choice. The f.o.c. indicate that both Poisson and gamma estimators are independent of scale, as only the constant, denoted β_0, is affected by the linear transformation according to $\tilde{\beta}_0(\lambda) = \beta_0 + Log\ \lambda$, such that $exp(\tilde{\beta}_0(\lambda)) = \lambda\ exp(\beta_0)$. This implies that $exp(X_i\ \tilde{\beta}(\lambda)) = \lambda\ exp(X_i\ \beta)\ \forall i$ with $\beta \equiv \tilde{\beta}(\lambda = 1)$, and the f.o.c. (7.1) and (7.3) are unaffected by scale:

$$\begin{cases} \tilde{\beta}_0(\lambda) = \beta_0 + Log\ \lambda \\ \tilde{\beta}_k(\lambda) = \beta_k \quad \forall k \neq 0 \end{cases} \Rightarrow \tag{7.4}$$

$$\text{PPML:} \quad \sum_i \big(\tilde{y}_i - exp(X_i\ \tilde{\beta})\big)\ X_i = 0 \Leftrightarrow \lambda \sum_i \big(y_i - exp(X_i\ \beta)\big)\ X_i = 0 \tag{7.5}$$

$$\text{GPML:} \quad \sum_i exp(-X_i\ \tilde{\beta})\ \big(\tilde{y}_i - exp(X_i\ \tilde{\beta})\big)\ X_i = 0 \Leftrightarrow \sum_i exp(-X_i\ \beta)\ \big(y_i - exp(X_i\ \beta)\big)\ X_i = 0 \tag{7.6}$$

In contrast, the f.o.c. for NB PML (eq. 7.2) might be sensitive to λ depending on how $\tilde{\alpha}(\lambda)$ adjusts. When $\tilde{y}$ is the dependent variable, the f.o.c. is:

$$\sum_i \big(1 + \tilde{\alpha}(\lambda)\ exp(X_i\ \tilde{\beta})\big)^{-1}\ \big(\tilde{y}_i - exp(X_i\ \tilde{\beta})\big)\ X_i = 0 \tag{7.7}$$

For $\tilde{\beta}$ to be unaffected (except for the constant) by scale, the following condition should hold:

$$\sum_i \big(1 + \tilde{\alpha}(\lambda)\ \lambda\ exp(X_i\ \beta)\big)^{-1}\ \lambda\big(y_i - exp(X_i\ \beta)\big)\ X_i = 0 \qquad (7.8)$$

The comparison of (7.2) and (7.8) implies that the condition under which the NB PML estimator is independent of λ is:

$$\textit{scale-invariance condition:}\quad \lambda\ \tilde{\alpha}(\lambda) = \alpha \quad \text{with} \quad \alpha \equiv \tilde{\alpha}(\lambda = 1).^{5} \qquad (7.9)$$

As one could expect the dispersion parameter $\tilde{\alpha}(\lambda)$ should vary with λ to ensure scale-invariance. Eq. (7.9) states that $\tilde{\alpha}(\lambda) = \frac{\alpha}{\lambda}$ compensates exactly for the change in the $\frac{Var(\tilde{y}|x)}{E(\tilde{y}|x)}$ ratio due to λ.

This scale-invariant condition is violated by the two-step NB PML (*i.e.* NB QGPML) estimators used in the literature. They consist of computing first a consistent estimator, such as PPML. α is then estimated from the first-step fitted values and residuals. In a second step, β is estimated from eq. (7.2) using this $\hat{\alpha}$.[6, 7] Gourieroux *et al.* (1984a) and Cameron and Trivedi (1986) obtain a consistent estimator of α from the following regression

5. Another way to see this is as follows. The NB PML assumption is $Var(\tilde{y}(\lambda)|X) = E[\tilde{y}(\lambda)|X] + \tilde{\alpha}(\lambda)\ E^2[\tilde{y}(\lambda)|X]$. Under the condition that $\tilde{\beta}$ is independent from λ (except $\tilde{\beta}_0$), this becomes: $\lambda^2\ Var[y|X] = \lambda\ E[y|X] + \lambda^2\ \tilde{\alpha}(\lambda)\ E^2[y|X] \Leftrightarrow Var[y|X] = 1/\lambda\Big(E[y|X] + \lambda\ \tilde{\alpha}(\lambda)\ E^2[y|X]\Big)$. Independence of this estimator (with robust standard errors) with respect to λ implies that $\lambda\ \tilde{\alpha}(\lambda) = \alpha$.

6. In Stata, one can use the *glm* command with *family(nbinomial a)* where $a = \hat{\alpha}$, which does not allow a to be negative. Some studies with a continuous dependent variable might have used the full NB likelihood estimator *nbreg* in Stata. While this is clearly inappropriate in such a case - we are grateful to one referee for highlighting this -, the NB ML estimator is also clearly scale-dependent because when y is a NB variable, $\tilde{y} = \lambda\ y$ is not due to the Γ function.

7. Estimating β and α jointly would require using the full NB log-likelihood (Wooldridge, 1999, p. 378), which is not suitable for non count data.

by OLS, where $\hat{y}_i$ denotes the first-step estimated values of y_i, and ϵ_i is a residual:

$$(y_i - \hat{y}_i)^2 - \hat{y}_i = \alpha \ \hat{y}_i{}^2 + \epsilon_i \tag{7.10}$$

This implies that:

$$\hat{\alpha} = \frac{\sum_i \left[(y_i - \hat{y}_i)^2 - \hat{y}_i\right] \ \hat{y}_i{}^2}{\sum_i \hat{y}_i{}^4} \tag{7.11}$$

This two-step estimator corresponds to the NB Quasi-Generalised PML estimator (NB QGPML) in Gourieroux *et al.* (1984a), called "two-step NB" by Wooldridge (1999) and denoted NB QGPML$_{OLS}$ herein. Indeed, Wooldridge (p. 380) argues that a more natural alternative is to estimate eq. (7.10) by weighted least squares (WLS) using $1/\sqrt{\hat{y}_i}$ as weights, which amounts to estimating (see also Wooldridge, 2002, p. 659):

$$\frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} - 1 = \alpha \ \hat{y}_i + \epsilon_i \tag{7.12}$$

This estimator is denoted here NB QGPML$_{WLS}$. Finally, Head *et al.* (2009) provide a Stata code for NB QGPML (NB QGPML$_{HMR}$) estimating α from eq. (7.10) above but using the geometric estimator in the first step (and the *glm* command in the second step, see footnote 6).

All these estimators depend on scale. They use the nominal variance assumption (Wooldridge, 1999, p. 378) by substracting $\hat{y}_i$ on the left-hand size of eq. (7.10) rather than estimating the unconstrained $(y_i - \hat{y}_i)^2 = a \ \hat{y}_i + b \ \hat{y}_i^2$ (see eq. (7.14) below). The formal proof is now given for NB QGPML$_{OLS}$, but is similar for the two other NB QGPML estimators just discussed. What happens to NB QGPML$_{OLS}$ when the linear transformation $\tilde{y} = \lambda\ y$ is used as the dependent variable? As shown before, the first-step PPML estimator is unaffected by scale, hence $\hat{\tilde{y}}_i(\lambda) = \lambda\ \hat{y}_i$.[8] It follows from eq. (7.11) that:

$$\begin{aligned}
\lambda\ \hat{\tilde{\alpha}}(\lambda) &= \lambda \frac{\sum_i [(\tilde{y}_i - \hat{\tilde{y}}_i)^2 - \hat{\tilde{y}}_i]\ \hat{\tilde{y}}_i^2}{\sum_i \hat{\tilde{y}}_i^4} = \lambda \frac{\sum_i [\lambda^2\ (y_i - \hat{y}_i)^2 - \lambda\ \hat{y}_i]\ \lambda^2\ \hat{y}_i{}^2}{\lambda^4\ \sum_i \hat{y}_i{}^4} \\
&= \frac{\sum_i [\lambda\ (y_i - \hat{y}_i)^2 - \hat{y}_i]\ \hat{y}_i{}^2}{\sum_i \hat{y}_i{}^4} = \hat{\alpha} + (\lambda - 1) \frac{\sum_i (y_i - \hat{y}_i)^2\ \hat{y}_i{}^2}{\sum_i \hat{y}_i{}^4}
\end{aligned} \tag{7.13}$$

which proves that:

(i) $\lambda\ \hat{\tilde{\alpha}}(\lambda) \neq \hat{\alpha}$ as soon as $\lambda \neq 1$. That is, the condition (eq. 7.9) under which the NB QGPML is independent of scale is violated;

(ii) when $\lambda \to +\infty$, $\hat{\tilde{\alpha}}(\lambda) \to \frac{\sum_i (y_i - \hat{y}_i)\ \hat{y}_i{}^2}{\sum_i \hat{y}_i{}^4} = \hat{\alpha} + \frac{\sum_i \hat{y}_i{}^3}{\sum_i \hat{y}_i{}^4}$, hence $\lambda\ \hat{\tilde{\alpha}}(\lambda) \to +\infty$ and NB QGPML $\to$ GPML following eq. (7.8);

(iii) when $\lambda \to 0$, $\lambda\ \hat{\tilde{\alpha}}(\lambda) \to \hat{\alpha} - \frac{\sum_i (y_i - \hat{y}_i)\ \hat{y}_i{}^2}{\sum_i \hat{y}_i{}^4} = -\frac{\sum_i \hat{y}_i{}^3}{\sum_i \hat{y}_i{}^4} < 0$. When the estimated value is constrained to be positive (such as NB QGPML$_{HMR}$ using the Stata *glm* command), which is often necessary to ensure that the estimated variance is positive, this tends to be binding and $\lambda\ \hat{\tilde{\alpha}}(\lambda) \to 0$, and NB QGPML $\to$ PPML.

7.2.2. A new scale-invariant NB QGPML estimator

The NB QGPML^{GLM} estimator that we now introduce overcomes this limitation. The idea is to use the analog of the Poisson GLM assumption (see Wooldridge, p. 357) for the PPML estimator and the procedure is the following. The first step estimates the following relation (which is also inspired by the conditional variance specification test proposed by

8. PPML point estimates are scale-invariant as discussed above, and estimated *robust* standard errors are also scale invariant, see *e.g.* Santos Silva and Tenreyro (2006). As will be obvious, the geometric mean used is the first step for NB QGPML$_{HMR}$ is also scale-dependent, which adds a second source of scale-dependence in that case.

Wooldridge, p. 372, eq. 8.43):

$$\hat{u_i}^2 = a\ \hat{y_i} + b\ \hat{y_i}^2 + \epsilon_i \tag{7.14}$$

This is equivalent to:

$$\hat{u_i}^2 = a\ (\hat{y_i} + \alpha\ \hat{y_i}^2) + \epsilon_i \quad \text{with} \quad \alpha = \frac{b}{a} \tag{7.15}$$

with $\hat{u_i} = y_i - \hat{y_i}$ being the first-step estimated residual. With robust standard errors, the second step computes the NB QGPML estimator using $\alpha = \frac{b}{a}$.

When $\tilde{y} = \lambda\ y$ is used as the dependent variable, $\hat{\tilde{u}}_i^2 = \lambda^2\ \hat{u_i}^2 = \lambda^2\ a\ \hat{y_i} + \lambda^2\ b\ \hat{y_i}^2 + \lambda^2 \epsilon_i = \lambda\ a\ \hat{\tilde{y}}_i + b\ \hat{\tilde{y}}_i^2 + \lambda^2\ \epsilon_i$, which implies $\tilde{a}(\lambda) = \lambda\ a$ and $\tilde{b}(\lambda) = b$, hence $\lambda\ \tilde{\alpha}(\lambda) = \lambda \frac{\tilde{b}(\lambda)}{\tilde{a}(\lambda)} = \lambda \frac{b}{\lambda\ a} = \frac{b}{a} = \alpha$ which ensures scale-invariance following eq. (7.9). This scale-invariant estimator is labelled NB QGPML$^{GLM}_{OLS}$. The corresponding WLS estimator (NB QGPML$^{GLM}_{WLS}$), which is more consistent with the GLM apparatus according to Wooldridge, is obtained by estimating $\alpha = \frac{b}{a}$ from: [9]

$$\frac{\hat{u_i}^2}{\hat{y_i}} = a + b\ \hat{y_i} + \epsilon_i \tag{7.16}$$

One important point refers to consistency. All estimators discussed above, including the scale-dependent NB QGPML estimators discussed in sub-section 7.2.1, are consistent among the class of exponential families. However, the main motivation for using NB

9. The Stata code for both NB QGPML$^{GLM}_{OLS}$ and NB QGPML$^{GLM}_{WLS}$ is available at https://sites.google.com/site/clementbosquet/. Using the consistent Poisson PML estimate in the first step, $\alpha = \frac{b}{a}$ is estimated from eq. (7.14) or (7.16), respectively.

estimators is to improve efficiency as the NB conditional variance assumption encompasses both the Poisson and gamma assumptions. Unfortunately, in contrast to the proposed NB QGPML^{GLM}, the advantage provided by the NB QGPML estimators that have been used so far is fake as they artificially cover the whole range of estimators between PPML and GPML depending only on the unit chosen for the dependent variable.

7.3. Monte-Carlo simulations

This section illustrates the scale-dependence of usual NB QGPML estimators through Monte-Carlo simulations when the dependent variable truly belongs to the class of exponential families. The simulations are based on the following assumptions.

$$X \sim \mathcal{N}(0,1) \tag{7.17}$$

$$E[y_i|x] = \mu(x_i\beta) = exp(\beta_0 + \beta_1 x_i) \tag{7.18}$$

$$\beta_0 = 0 \qquad \text{and} \qquad \beta_1 = 1 \tag{7.19}$$

$$y_i = \mu(x_i\beta)\eta_i \tag{7.20}$$

where η_i is a log normal random variable such that $E[\eta_i] = 1$ and $V(\eta_i) = \sigma_i^2$ given by

$$\sigma_i^2 = \mu(x_i\beta)^{-1} + \alpha \tag{7.21}$$

Hence, $V(y_i|x) = \mu(x_i\beta)^2\ \sigma_i^2 = E[y_i|x] + \alpha\ E[y_i|x]^2$.

This is the standard case where the NB PML estimator is optimal. We use $\alpha = 0.5$

and a sample size of 1,000. To illustrate the artificial scale-dependence (using $\tilde{y}_i = \lambda\ y_i$ for various λ) of the usual NB QGPML estimators, Table 7.1 focuses on one single draw and provides the estimated values across all the estimators discussed in the preceding section for 3 different values of λ. Panels A, B and C present the results obtained by NB QGPML$_{OLS}$, NB QGPML$_{WLS}$ and NB QGPML$_{HMR}$, respectively. All panels also include the scale-invariant PPML, GPML, NB QGPML$^{GLM}_{OLS}$ and NB QGPML$^{GLM}_{WLS}$. [10]

As expected the usual NB QGPML estimators are scale-dependent converging towards either PPML or GPML depending on λ, with $\tilde{\alpha}(\lambda)$ behaving as described and the estimated constant being close numerically to $\beta_0 + Log(\lambda)$. Second, estimates are not significantly different across estimators.

Table 7.2 presents the results based on 10,000 replicas, with Panel A providing the estimated $\lambda\ \tilde{\alpha}(\lambda) = \alpha$ for the scale-dependent NB QGPML and $\lambda\ \tilde{a}(\lambda) = a$ and $\tilde{b}(\lambda) = b$ for NB QGPML^{GLM}. All three parameters ($\tilde{\alpha}$, $\tilde{a}$ and $\tilde{b}$) are constrained to be positive in the second step. [11] When $\lambda = 1/1000$ the unconstrained $\tilde{\alpha}$ are negative for all replicas whether NB QGPML$_{OLS}$, NB QGPML$_{WLS}$ or NB QGPML$_{HMR}$ is used. In these cases, the second-step estimates coincide with PPML. Scale-dependence is also illustrated when $\lambda = 1000$: in that case $\lambda\ \tilde{\alpha}(\lambda)$ grows very large as shown in the proof given in sub-section 7.2.1. For example, with NB QGPML$_{WLS}$, $\lambda\ \tilde{\alpha}(\lambda)$ is equal to about 643 on average and estimates tend towards GPML. Panel B shows the results for the average bias of the parameters of interests β_0 and β_1 with their standard errors. As expected all estimators are consistent. Beyond scale-invariance, NB QGPML^{GLM} estimators perform well also, and NB PML$^{GLM}_{WLS}$

10. Sub-section 7.2.2 has proved that our NB QGPML^{GLM} estimators are scale-invariant. Of course, we checked that this theoretical property is confirmed by the empirical analysis!

11. Otherwise, it is very often the case that the estimate of the conditional variance becomes negative for at least a few observations, which prevents the second-step estimation.

Table 7.1: Scale-dependence of the usual Negative-Binomial Estimator: 1 single draw $\alpha = 0.5$

	PPML	NB QGPML			GPML	NB QGPML^{GLM}	
		$\lambda = \frac{1}{1000}$	$\lambda = 1$	$\lambda = 1000$			
Panel A: NB QGPML$_{OLS}$							
	PPML	NB$_{OLS}$	NB$_{OLS}$	NB$_{OLS}$	GPML	NB$^{GLM}_{OLS}$	NB$^{GLM}_{WLS}$
β_1 (x)	0.968^a	0.968^a	1.001^a	1.051^a	1.052^a	0.969^a	1.017^a
	(0.037)	(0.037)	(0.036)	(0.040)	(0.040)	(0.037)	(0.034)
$\beta_0 = constant - Log(\lambda)$	-0.042	-0.042	-0.063	-0.089^b	-0.089^b	-0.043	-0.073^c
	(0.040)	(0.040)	(0.039)	(0.038)	(0.038)	(0.040)	(0.038)
Observations	1000	1000	1000	1000	1000	1000	1000
$\tilde{\alpha}(\lambda)$		-72.8735	0.1297	0.2027			
$\lambda\ \tilde{\alpha}(\lambda)$		-0.0729	0.1297	202.6711		0.0014	0.4300
$\lambda\ \tilde{a}(\lambda)$						2.7241	0.8349
$\tilde{b}(\lambda)$						0.0037	0.3590
Panel B: NB QGPML$_{WLS}$							
	PPML	NB$_{WLS}$	NB$_{WLS}$	NB$_{WLS}$	GPML	NB$^{GLM}_{OLS}$	NB$^{GLM}_{WLS}$
β_1 (x)	0.968^a	0.968^a	1.013^a	1.052^a	1.052^a	0.969^a	1.017^a
	(0.037)	(0.037)	(0.034)	(0.040)	(0.040)	(0.037)	(0.034)
$\beta_0 = constant - Log(\lambda)$	-0.042	-0.042	-0.070^c	-0.089^b	-0.089^b	-0.043	-0.073^c
	(0.040)	(0.040)	(0.038)	(0.038)	(0.038)	(0.040)	(0.038)
Observations	1000	1000	1000	1000	1000	1000	1000
$\tilde{\alpha}(\lambda)$		-269.1896	0.3145	0.5840			
$\lambda\ \tilde{\alpha}(\lambda)$		-0.2692	0.3145	583.9846		0.0014	0.4300
$\lambda\ \tilde{a}(\lambda)$						2.7241	0.8349
$\tilde{b}(\lambda)$						0.0037	0.3590
Panel C: NB QGPML$_{HMR}$							
	PPML	NB$_{HMR}$	NB$_{HMR}$	NB$_{HMR}$	GPML	NB$^{GLM}_{OLS}$	NB$^{GLM}_{WLS}$
β_1 (x)	0.968^a	0.968^a	1.003^a	1.051^a	1.052^a	0.969^a	1.017^a
	(0.037)	(0.037)	(0.035)	(0.040)	(0.040)	(0.037)	(0.034)
$\beta_0 = constant - Log(\lambda)$	-0.042	-0.042	-0.064^c	-0.089^b	-0.089^b	-0.043	-0.073^c
	(0.040)	(0.040)	(0.039)	(0.038)	(0.038)	(0.040)	(0.038)
Observations	1000	1000	1000	1000	1000	1000	1000
$\tilde{\alpha}(\lambda)$		-72.7308	0.1551	0.2302			
$\lambda\ \tilde{\alpha}(\lambda)$		-0.0727	0.1551	230.1770		0.0014	0.4300
$\lambda\ \tilde{a}(\lambda)$						2.7241	0.8349
$\tilde{b}(\lambda)$						0.0037	0.3590

Standard error between parentheses. ^{a}, ^{b}, ^{c} Significant at the 1%, 5% and 10% level, respectively. PML = Pseudo-Maximum Likelihood, QGPML = Quasi-Generalised PML, PPML = Poisson PML, NB = Negative Binomial, GPML = gamma PML.

seems to outperform NB QGPML$^{GLM}_{OLS}$, as suggested by Wooldridge (1999). For the former, a is equal on average to 0.97 close to the true value of 1 while b averages to 0.44 close to

0.5. Moreover, the NB QGPML$_{WLS}$ procedure leads to GPML ($a = 0$) for only 2% of the case and PPML ($b = 0$) for 1%, compared with 12% and 23%, respectively, for NB QGPML$_{OLS}$. In sum, Table 7.2 illustrate the four expected results: scale-dependence of the usual NB QGPML, superiority of NB QGPML^{GLM} (especially NB QGPML$^{GLM}_{WLS}$) in the sense of scale-invariance (and efficiency), and consistency of all estimators.[12]

7.4. Application to the trade gravity equation

7.4.1. Data

Trade flow data are taken from the IMF Direction Of Trade Statistics (DOTS) database. For the year 2000, there are 21,543 non-zero flows between 196 trading partners.[13] The geographical variables (distance between countries, common border, common language and colonial linkage dummies) are provided by the CEPII database[14], and the FTA data are based on Fontagné and Zignago (2007) who improve those used by Baier and Bergstrand (2007).

7.4.2. Specification

The bilateral trade equation is estimated following Anderson and van Wincoop (2003) according to:

$$x_{ij} = exp(\beta_0 + \beta_1\ Log\ d_{ij} + \beta_2\ B_{ij} + \beta_3\ L_{ij} + \beta_4\ C_{ij} + \beta_5\ FTA_{ij} + FX_i + FM_j)\eta_{ij} \quad (7.22)$$

12. As suggested by referees, we tested different settings such as $\alpha = 2$ and / or a gamma random variable for η (instead of log normal) and obtained similar results.

13. Focusing on non-zero flows is sufficient for illustration purposes. Including zero flows or focusing on other years unsurprisingly leads to the same conclusion as the proof in section 7.2 is general.

14. http://www.cepii.fr/anglaisgraph/bdd/distances.htm, Centre d'Etudes Prospectives et d'Informations Internationales.

Table 7.2: Scale-dependence of the usual Negative-Binomial Estimator, simulations, 10,000 replicas $\alpha = 0.5$

Panel A

	$\lambda\ \tilde{\alpha}(\lambda)$		$\lambda\ \tilde{a}(\lambda)$		$b(\lambda)$	
Estimator	Mean	$\% > 0$ *	Mean	$\% > 0$ *	Mean	$\% > 0$ *
NB_{OLS} ($\lambda = \frac{1}{1000}$)	-0.05706	0%	.	.	.	.
NB_{WLS} ($\lambda = \frac{1}{1000}$)	-0.22819	0%	.	.	.	.
NB_{HMR} ($\lambda = \frac{1}{1000}$)	-0.05713	0%	.	.	.	.
NB_{OLS} ($\lambda = 1$)	0.26494	100%	.	.	.	.
NB_{WLS} ($\lambda = 1$)	0.41417	100%	.	.	.	.
NB_{HMR} ($\lambda = 1$)	0.29615	100%	.	.	.	.
NB_{OLS} ($\lambda = 1000$)	324.32885	100%	.	.	.	.
NB_{WLS} ($\lambda = 1000$)	642.81816	100%	.	.	.	.
NNB_{HMR} ($\lambda = 1000$)	350.30616	100%	.	.	.	.
NB^{GLM}_{OLS}	.	.	2.50537	87%	0.32150	77%
NB^{GLM}_{WLS}	.	.	0.97176	98%	0.43647	99%

Panel B

	β_0		β_1	
Estimator	Bias	S.E.	Bias	S.E.
PPML	-0.00055	0.04877	-0.00199	0.05348
NB_{OLS} ($\lambda = \frac{1}{1000}$)	-0.00055	0.04877	-0.00199	0.05348
NB_{WLS} ($\lambda = \frac{1}{1000}$)	-0.00055	0.04877	-0.00199	0.05348
NB_{HMR} ($\lambda = \frac{1}{1000}$)	-0.00055	0.04877	-0.00199	0.05348
NB_{OLS} ($\lambda = 1$)	0.00055	0.04361	-0.00300	0.04175
NB_{WLS} ($\lambda = 1$)	-0.00037	0.04337	-0.00143	0.04085
NB_{HMR} ($\lambda = 1$)	0.00102	0.04360	-0.00381	0.04163
NB_{OLS} ($\lambda = 1000$)	-0.00200	0.04550	0.00379	0.05354
NB_{WLS} ($\lambda = 1000$)	-0.00204	0.04554	0.00394	0.05382
NB_{HMR} ($\lambda = 1000$)	-0.00198	0.04551	0.00372	0.05361
GPML	-0.00207	0.04556	0.00406	0.05403
NB^{GLM}_{OLS}	0.00212	0.04526	-0.00501	0.04706
NB^{GLM}_{WLS}	-0.00047	0.04336	-0.00115	0.04103

PML = Pseudo-Maximum Likelihood, QGPML = Quasi-Generalised PML, PPML = Poisson PML, NB = Negative Binomial QGPML, GPML = gamma PML.
(*) These columns indicate the percentage of occurrences for which the estimated values are positive.

where x_{ij} is the value of export from country i to country j, FX_i and FM_j are exporting and importing countries fixed effects, respectively. B_{ij}, L_{ij} and C_{ij} are the traditionnal control covariates: common border, common official language and colonial linkage dummies, respectively. The η_{ij} are the multiplicative error terms of the nonlinear estimates.

Following Anderson and van Wincoop (2003), importer and exporter fixed effects are used to control for multilateral resistance terms as well as for the income levels of both importers and exporters.

7.4.3. Results

Table 7.3 shows the estimates of equation (7.22) from the PPML, GPML and NB QGPML$_{WLS}^{GLM}$ estimators, in the first, penultimate and last columns, respectively. The columns in between report scale-dependent NB QGPML$_{OLS}$ estimates based on different unit values for trade flows. Unreported NB QGPML$_{WLS}$ and NB QGPML$_{HMR}$ estimates yield similar results.

Table 7.3: Scale-dependence of the Negative-Binomial Estimator: gravity equation, 2000

	PPML	NB QGPML$_{OLS}$				GPML	NB QGPML$_{WLS}^{GLM}$
		Tr. USD	B. USD	M. USD	Th. USD		
Distance	-0.606^a	-0.606^a	-0.650^a	-1.125^a	-1.254^a	-1.231^a	-0.898^a
	(0.032)	(0.032)	(0.035)	(0.024)	(0.024)	(0.024)	(0.028)
Contiguity	0.644^a	0.644^a	0.590^a	0.685^a	0.926^a	0.959^a	0.450^a
	(0.099)	(0.099)	(0.099)	(0.079)	(0.089)	(0.090)	(0.086)
Common-language	0.154^c	0.154^c	0.167^b	0.395^a	0.533^a	0.541^a	0.181^a
	(0.082)	(0.082)	(0.081)	(0.048)	(0.050)	(0.051)	(0.058)
Colonial-tie	0.175^c	0.175^c	0.275^b	1.079^a	1.289^a	1.289^a	0.821^a
	(0.103)	(0.103)	(0.118)	(0.076)	(0.096)	(0.100)	(0.089)
Free-trade agreement	0.479^a	0.479^a	0.425^a	0.137^b	0.210^a	0.250^a	0.213^a
	(0.088)	(0.088)	(0.083)	(0.056)	(0.061)	(0.061)	(0.062)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21543	21543	21543	21543	21543	21543	21543
$\tilde{\alpha}(\lambda)$		-6.3957	0.0151	0.0215	0.0215		
$\lambda\ \tilde{\alpha}(\lambda)$		-0.0064	0.0151	21.4699	21476.27		0.5534
$\lambda\ \tilde{a}(\lambda)$							0.1691
$\tilde{b}(\lambda)$							0.0936

Standard error between brackets. [a], [b], [c] Significant at the 1%, 5% and 10% level, respectively. PML = Pseudo-Maximum Likelihood, QGPML = Quasi-Generalised PML, PPML = Poisson PML, NB = Negative Binomial GQPML, GPML = gamma PML ; USD = United States Dollars, Tr.=Trillions, B.=Billions, M.=Millions, Th.=Thousands. Fixed effects are importer and exporter country fixed effects.

When flows are measured with a very large unit (billions of US$), which means that flow values are small, NB QGPML and PPML estimates are close to each other. They coincide exactly when trade flows are measured in trillions of US$. At the other extreme, when flows take very large values (*i.e.* when the unit is small such as thousands of US$), the NB QGPML are very close to the GPML estimates. This illustrates that the usual NB QGPML estimator depends arbitrarily on the unit choice of the dependent variable, with $\hat{\hat{\alpha}}(\lambda)$ behaving as expected. In contrast, our NB QGPML$_{WLS}^{GLM}$ estimator is scale-invariant and might improve upon both PPML and GPML. With NB QGPML$_{WLS}^{GLM}$, the point estimate of the distance coefficient is -0.898 compared with -0.606 for PPML and -1.231 for GPML.

As an aside, this empirical exercise shows that, when applied to the gravity equation of trade flows, PPML, NB QGPML^{GLM} and GPML lead to point estimates that differ significantly. This is problematic because, beyond efficiency issues, the underlying assumption is that these estimators are mutually consistent.

7.5. Conclusion

The NB QGPML is being increasingly used as a way to improve efficiency relative to either the PPML or GPML estimators. This note shows that the NB QGPML estimators which have been used so far do not achieve this when the unit choice of the dependent variable is arbitrary, as in most models with a continuous dependent variable, because, by using the nominal variance assumption (Wooldridge, 1999), they artificially depend on scale in such cases. A new NB QGPML estimator, which relies on the GLM variance assumption, is introduced and overcomes this limitation.

Conclusion générale

Ainsi, dans la première partie de cette thèse a d'abord été analysé l'impact des choix méthodologiques sur les classements des universités et centres de recherche en économie en France en 2008. Nous avons montré que la hiérarchie des institutions évaluée avec les citations Google Scholar était très proche de celle obtenue plus traditionnellement avec les publications Econlit. Utiliser la qualité des journaux comme prédicteur de la qualité des publications, en tous les cas du nombre de citations qu'elles reçoivent, semble donc constituer une stratégie raisonnable. Pour quelques institutions, dont le cœur d'activité n'est pas l'économie, l'écart entre les approches par citations et par publications est plus important, ce qui nous fait considérer que Google Scholar est un outil complémentaire d'Econlit. Prendre ou non en compte le nombre d'auteurs par article, ramener les citations par entrée à celles en ayant au moins une, considérer certains indicateurs ou certaines périodes particulières n'a qu'un impact très marginal sur le panorama de la recherche française en économie que nous avons dessiné.

Puis dans le deuxième chapitre, ont été étudiés les déterminants des stocks individuels de publications et de citations des chercheurs. Nous avons mis à jour des rendements d'échelle croissants de la quantité et de la taille des réseaux de co-auteurs sur la qualité de la production scientifique ainsi que des effets de réseaux importants dans la diffusion de la connaissance. En effet, les chercheurs les plus cités sont naturellement ceux qui ont le plus publier (en quantité et en qualité) mais ceux qui ont travaillé avec des équipes de co-auteurs plus large et qui ont un réseau de stock de co-auteurs différents plus important sont encore davantage cités.

Ensuite, dans la deuxième partie de cette thèse, nous avons cherché à établir des relations causales en économie de la science avec des applications en économie géographique et en économie du travail. En effet, le troisième chapitre étudie les déterminants individuels et locaux de la productivité des chercheurs quand le quatrième chapitre s'intéresse aux différentiels de promotion entre hommes et femmes sur le marché du travail académique. Contrairement à la plupart des autres articles de la littérature, nous avons conclu à la présence d'effets de pairs dans la recherche académique en montrant que la localisation des chercheurs avait un impact sur leur productivité scientifique. Les universités qui génèrent le plus d'externalités positives sont celles dans lesquelles les chercheurs sont homogènes en terme de performances, la diversité thématique est grande, et, dans une moindre mesure, dans lesquelles il y a beaucoup de chercheurs, plus de femmes, de chercheurs âgés, de stars et dans lesquelles les chercheurs sont connectés à des co-auteurs à l'étranger.

Dans le quatrième chapitre, nous avons étudié l'écart de poste atteint entre hommes et femmes sur le marché du travail académique, tant en termes de promotion que de qualité scientifique du lieu de travail. Nous avons utilisé les particularités du système universitaire français pour conclure à l'absence de discrimination contre les femmes dans le processus de promotion. Si elles sont moins souvent Professeurs des universités (par opposition à Maître de conférences) ou Directrices de recherche (par opposition à Chargé de recherche) que les hommes, ce n'est pas non plus parce que le coût de la promotion (mobilité pour les universitaires) est plus important pour elles, ni parce qu'elles ont des préférences différentes concernant le salaire et le prestige des institutions dans lesquelles elles travaillent. Une explication possible, mais non testée, serait donc que les femmes s'autocensurent ou exercent moins d'efforts que les hommes pendant les processus de promotion.

Enfin, la troisième partie s'inscrit dans une littérature différente consacrée au choix des estimateurs pour les équations de gravité appliquées au commerce international. Elle commence, dans le cinquième chapitre, par relativiser l'article qui fait référence en la matière et qui préconise d'estimer les modèles économétriques de manière générale et les équations de gravité en particulier directement en niveau puisque leurs corollaires log-linéarisés seraient biaisés. En effet, cette analyse repose sur une hypothèse qu'il est nécessaire de discuter et ce cinquième chapitre montre que si c'est l'hypothèse concurrente qui est la bonne, alors ce sont les estimations en niveau qui sont biaisés.

Puis, le sixième chapitre étudie le "paradoxe de la distance" (augmentation de la valeur absolue du coefficient associé à la distance dans les équations de gravité) en comparant différents estimateurs possibles. Nous avons montré que l'utilisation de l'estimateur de pseudo-maximum de vraisemblance à partir d'une loi de Poisson au modèle gravitaire en niveau conduit à l'obtention d'une élasticité du commerce à la distance constante dans le temps. L'écart grandissant entre les estimations par moindres carrés ordinaires du modèle log-linéarisé et par pseudo-maximum de vraisemblance à partir d'une loi de Poisson du modèle en niveau semble être dû à l'hétérogénéité croissante des flux de commerce.

Finalement, le septième chapitre montre que les estimateurs de pseudo-maximum de vraisemblance à partir d'une loi binomiale négative qui ont été utilisés dans la littérature dépendent de l'échelle/unité de mesure de la variable dépendante. Ainsi, les paramètres estimés d'une équation de gravité vont être différents selon que les flux de commerce soient mesurés en milliers d'euros ou en millions de dollars par exemple. Nous présentons une nouvelle version en deux étapes de cet estimateur qui ne possède pas cette propriété.

Les prolongements naturels de cette thèse s'inscrivent dans la suite de la partie principale consacrée à l'économie de la science. A l'aide d'une approche issue de l'économie géographique empirique, nous avons conclu à la présence d'effets de pairs dans la recherche, contrairement à un certain nombre d'autres papiers de la littérature. Bramoullé et al. (2009) ont montré qu'il était possible d'utiliser les réseaux sociaux pour identifier les effets de pairs quand le problème de "réflexion" mis à jour par Manski (1993) était à l'œuvre. Certains auteurs ont alors commencé à utiliser les réseaux sociaux afin d'identifier des effets de pairs en économie de l'éducation (Calvó-Armengol, Patacchini et Zenou, 2009) et des effets de pairs dans la recherche (Ductor, 2011). La limite principale de ce type d'approche est l'endogénéité liée à la formation des réseaux : les liens sociaux ne sont pas formés au hasard. On peut imaginer que cette limite n'est pas très importante en économie de l'éducation dans la mesure où la formation des liens d'amitié à l'école primaire ou au collège ne semble *a priori* pas découler d'un processus de maximisation utilitariste très puissant de la part des élèves. Néanmoins, cette limite paraît beaucoup plus génante si l'on veut étudier les effets de pairs par les réseaux de co-auteurs dans la recherche académique. Je souhaite donc mener un projet de recherche en deux étapes dont la première sera l'étude de la formation des réseaux de co-autorat. Seront étudiés les déterminants des choix de collaboration scientifique, y compris les déterminants géographiques qui sont absents de la littérature existante (Fafchamps, Goyal et van der Leij, 2010). Dans la deuxième étape, je pourrais alors étudier les effets de pairs dans la recherche conditionnellement à cette formation endogène des réseaux de co-autorat.

Parmi les prolongements possibles à la troisième partie de cette thèse, il serait intéressant et potentiellement fort utile de construire un test statistique permettant de discriminer entre une estimation par moindres carrés ordinaires d'un modèle log-linéarisé et une estimation du modèle en niveau par un estimateur de pseudo-maximum de vraisemblance à partir d'une loi de Poisson dans la mesure où le test P_ϵ de MacKinnon, White et Davidson (1983) n'est applicable que dans les cas où les résidus des deux modèles alternatifs sont identiquement distribués et suivent une loi Normale, comme illustré dans le cinquième chapitre de cette thèse.

Bibliographie

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999), 'High Wage Workers and High Wage Firms', *Econometrica* **67**(2), 251–333.

Acs, Z. J., Audretsch, D. B. and Feldman, M. P. (1991), 'Real Effects of Academic Research : A Comment', *American Economic Review* **82**(1), 363–367.

Agrawal, A. and Goldfarb, A. (2008), 'Restructuring Research : Communication Costs and the Democratization of University Innovation', *American Economic Review* **98**(4), 1578–1590.

Anderson, J. and van Wincoop, E. (2003), 'Gravity with Gravitas : A Solution to the Border Puzzle', *American Economic Review* **93**(1), 170–192.

Arzaghi, M. and Henderson, J. V. (2008), 'Networking off Madison Avenue', *Review of Economic Studies* **75**(4), 1011–1038.

Auriol, E., Friebel, G. and Lammers, F. (2012), The Firm as the Locus of Social Comparisons : Internal Labor Markets versus Up-or-Out, Discussion Paper 8831, Centre for Economic Policy Research.

Azoulay, P., Graff Zivin, J. S. and Wang, J. (2010), 'Superstar Extinction', *Quarterly Journal of Economics* **125**(2), 549–589.

Baier, S. L. and Bergstrand, J. H. (2007), 'Do Free Trade Agreements Actually Increase Members International Trade ?', *Journal of International Economics* **71**(1), 72–95.

Baldwin, R. and Taglioni, D. (2006), 'Gravity for Dummies and Dummies for Gravity Equations', *CEPR Discussion Paper* **5850**.

Berthelon, M. and Freund, C. (2008), 'On the Conservation of Distance in International Trade', *Journal of International Economics* **75**(2), 310–320.

Besley, T. and Ghatak, M. (2008), 'Social Incentives', *American Economic Review* **98**(2), 206–211.

Blackaby, D., Booth, A. L. and Frank, J. (2005), 'Outside Offers and the Gender Pay Gap : Empirical Evidence from the UK Academic Labour Market', *Economic Journal* **115**, 81–107.

Blau, F. and Kahn, L. M. (2000), 'Gender Differences in Pay', *Journal of Economic Perspectives* **14**(4), 75–99.

Bosquet, C. and Boulhol, H. (2010), 'Scale-dependence of the Negative Binomial Pseudo-Maximum Likelihood Estimator', *GREQAM Working Paper* **2010-39**.

Bosquet, C. and Combes, P.-P. (2011), *Comparaison des mesures Econlit et Google Scholar de la production de recherche en économie en France en 2008*, Direction Générale de la Recherche et de l'Innovation (DGRI) du Ministère de l'Enseignement Supérieur et de la Recherche.

Bosquet, C. and Combes, P.-P. (2012), 'Do Large Departments Make Academics More Productive ? Agglomeration and Peer Effects in Research', *Mimeo Greqam* .

Bosquet, C., Combes, P.-P. and Linnemer, L. (2010), *La publication d'articles de recherche en économie en France en 2008. Disparités actuelles et évolutions depuis 1998*, Direction Générale de la Recherche et de l'Innovation (DGRI) du Ministère de l'Enseignement Supérieur et de la Recherche.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009), 'Identification of Peer Effects Through Social Networks', *Journal of Econometrics* **150**(1), 41–55.

Briant, A., Combes, P.-P. and Lafourcade, M. (2009), 'Product Complexity, Quality of Institutions and the Pro-Trade Effect of Immigrants', *CEPR Discussion Paper* (7192).

Buch, C. M., Kleinert, J. and Toubal, F. (2004), 'The Distance Puzzle : On the Interpretation of the Distance Coefficient in Gravity Equations', *Economics Letters* **83**(3), 293–298.

Burger, M., van Oort, F. and Linders, G.-J. (2009), 'On the Specification of the Gravity Model of Trade : Zeros, Excess Zeros and Zero-inflated Estimation', *Spatial Economic Analysis* **4**(2), 167–190.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009), 'Peer Effects and Social Networks in Education', *Review of Economic Studies* **76**(4), 1239–1267.

Cameron, A. C. and Trivedi, P. K. (1986), 'Econometric Models Based on Count data : Comparisons and Applications of some Estimators and Tests', *Journal of Applied Econometrics* **1**(1), 29–53.

Coe, D. T., Subramanian, A. and Tamirisa, N. (2007), 'The Missing Globalization Puzzle : Evidence of the Declining Importance of Distance', *IMF Staff Papers* **54**(1).

Cole, J. and Cole, S. (1973), *Social Stratification in Science*, Chicago : The Univeristy of Chicago Press.

Combes, P.-P., Duranton, G. and Gobillon, L. (2008), 'Spatial Wage Disparities : Sorting Matters !', *Journal of Urban Economics* **63**(2), 723–742.

Combes, P.-P. and Linnemer, L. (2001), 'La publication d'articles de recherche en économie en France', *Annales d'Économie et de Statistiques* **62**, 5–47.

Combes, P.-P. and Linnemer, L. (2003*a*), 'L'impact international des articles de recherche français en économie', *Revue Économique* **54**, 181–217.

Combes, P.-P. and Linnemer, L. (2003*b*), 'Where are the Economists who Publish ? Publication Concentration and Rankings in Europe Based on Cumulative Publications', *Journal of the European Economic Association* **1**(6), 1250–1308.

Combes, P.-P. and Linnemer, L. (2010), 'Inferring Missing Citations. A Quantitative Multi-Criteria Ranking of all Journals in Economics', *GREQAM Working Paper* (2010-25).

Combes, P.-P., Linnemer, L. and Visser, M. (2008), 'Publish or Peer-Rich ? The Role of Skills and Networks in Hiring Economics Professors', *Labour Economics* **15**(3), 423–441.

Combes, P.-P., Mayer, T. and Thisse, J.-F. (2008), *Economic Geography : The integration of Regions and Nations*, Princeton University Press, New Jersey.

Davidson, R. and MacKinnon, J. G. (1981), 'Several Tests for Model Specification in the Presence of Alternative Hypotheses', *Econometrica* **49**(3), 781–793.

Deardoff, A. (1998), *Determinants of Bilateral Trade : Does Gravity Work in a Neoclassical World ?*, in Jeffrey Frankel (Ed.), *The Regionalization of the World Economy* ; Chicago : University of Chicago Press.

Diamond, A. M. (1986), 'What is a Citation Worth ?', *Journal of Human Resources* **21**(2), 200–215.

Disdier, A.-C. and Head, K. (2008), 'The Puzzling Persistence of the Distance Effect on Bilateral Trade', *The Review of Economics and Statistics* **90**(1), 37–48.

Dolado, J. J., Felgueroso, F. and Almunia, M. (2008), 'Do Men and Women-Economists Choose the Same Research Fields ? : Evidence from Top-50 Departments', *Fedea Documento de Trabajo* (2008-15).

Dubois, P., Rochet, J.-C. and Schlenker, J.-M. (2011), Productivity and Mobility in Academic Research : Evidence from Mathematicians. Processed, Toulouse School of Economics.

Ductor, L. (2011), Does Co-authorship Lead to Higher Academic Productivity ? Processed, Universidad de Alicante.

Ductor, L., Fafchamps, M., Goyal, S. and van der Leij, M. J. (2011), Social Networks and Research Output. Processed, Oxford University.

Egghe, L. (2006), 'Theory and Practise of the G-index', *Scientometrics* **69**(1), 131–152.

Fafchamps, M., Goyal, S. and van der Leij, M. J. (2010), 'Matching and Network Effects', *Journal of the European Economic Association* **8**(1), 203–231.

Farber, S. (1977), 'The Earnings and Promotion of Women Faculty : Comment', *American Economic Review* **67**(2), 199–206.

Felbermayr, G. J. and Kohler, W. (2006), 'Exploring the Intensive and Extensive Margins of World Trade', *Review of World Economics* **142**(4), 642–674.

Fontagné, L. and Zignago, S. (2007), 'A Re-evaluation of the Impact of Regional Agreements on Trade Patterns', *Economie Internationale* **109**(1), 31–51.

Garita, G. and van Marrewijk, C. (2008), 'Countries of a Feather Flock Together – Mergers and Acquisitions in the Global Economy', *Tinbergen Institute Discussion Paper* **067**(2).

Gaspar, J. and Glaeser, E. L. (1998), 'Information Technology and the Future of Cities', *Journal of Urban Economics* **43**(1), 136–156.

Ginther, D. K. and Hayes, K. J. (1999), 'Gender Differences in Salary and Promotion for Faculty in the Humanities', *American Economic Review* **89**(2), 397–402.

Ginther, D. K. and Kahn, S. (2004), 'Women in Economics : Moving Up or Falling Off the Academic Career Ladder ?', *Journal of Economic Perspectives* **18**(3), 193–214.

Gneezy, U., Niederle, M. and Rustichini, A. (2003), 'Performance in Competitive Environments : Gender Differences', *Quarterly Journal of Economics* **118**(3), 1049–1074.

Gobillon, L. (2004), The Estimation of Cluster Effects in Linear Panel Models. Processed, INED.

Goldin, C. and Rouse, C. (2000), 'Orchestrating Impartiality : The Impact of "Blind" Auditions on Female Musicians', *American Economic Review* **90**(4), 715–742.

Golub, S. and Tomasik, B. (2008), 'Measures of International Transport Cost for OECD Countries', *OECD Economics Department Working Papers* (609).

Gourieroux, C., Monfort, A. and Trognon, A. (1984*a*), 'Pseudo Maximum Likelihood Methods : Applications to Poisson Models', *Econometrica* **52**(3), 701–720.

Gourieroux, C., Monfort, A. and Trognon, A. (1984*b*), 'Pseudo Maximum Likelihood Methods : Theory', *Econometrica* **52**(3), 681–700.

Goyal, S., van der Leij, M. J. and Moraga-González, J. L. (2006), 'Economics : An Emerging Small World', *Journal of Political Economy* **114**(2), 403–412.

Hamermesh, D. S., Johnson, G. E. and Weisbrod, B. A. (1982), 'Scholarship, Citations and Salaries : Economic Rewards in Economics', *Southern Economic Journal* **49**(2), 472–481.

Hansen, W. L., Weisbrod, B. A. and Strauss, R. P. (1978), 'Modeling the Earnings and Research Productivity of Academic Economists', *Journal of Political Economy* **86**(4), 729–741.

Harzing, A.-W. (2007), *Publish or Perish*, http ://www.harzing.com/pop.htm.

Head, K., Mayer, T. and Ries, J. (2009), 'How Remote is the Offshoring Threat ?', *European Economic Review* **53**(4), 429–444.

Hellerstein, J. K., Neumark, D. and Troske, K. R. (1999), 'Wages, Productivity, and Worker Characteristics : Evidence from Plant-level Production Functions and Wage Equations', *Journal of Labour Economics* **17**(3), 409–446.

Helpman, E., Melitz, M. and Rubinstein, Y. (2008), 'Estimating trade flows : Trading partners and trading volumes', *The Quaterly Journal of Economics* **123**(2), 441–487.

Hirsh, J. (2005), 'An Index to Quantify an Individual's Scientific Research Output', *Proceedings of the National Academy of Sciences* **102**, 16569–16572.

Hollis, A. (2001), 'Co-authorship and the Output of Academic Economists', *Labour Economics* **8**(4), 503 – 530.

Hummels, D. (2007), 'Transportation Costs and International Trade in the Second Era of Globalization', *Journal of Economic Perspectives* **21**(3), 131–154.

Jackson, M. O. (2011), An Overview of Social Networks and Economic Applications, *in* A. B. Jess Benhabib and M. O. Jackson, eds, 'Handbook of Social Economics', Vol. 1, North-Holland, Amsterdam, pp. 511–585.

Jacobs, J. (1969), *The Economy of Cities*, Random House, New York.

Jaffe, A. B. (1989), 'Real Effects of Academic Research', *American Economic Review* **79**(5), 957–970.

Johnson, G. E. and Stafford, F. P. (1974), 'Lifetime Earnings in a Professional Labor Market : Academic Economists', *Journal of Political Economy* **82**(3), 549–569.

Kahn, S. (1995), 'Gender Differences in Academic Career Paths of Economists', *Journal of Economic Perspectives* **9**(4), 193–204.

Katz, D. A. (1973), 'Faculty Salaries, Rates of Promotion and Productivity at a Large University', *American Economic Review* **63**(3), 469–477.

Kenny, L. W. and Studley, R. E. (1995), 'Economists' Salaries and Lifetime Productivity', *Southern Economic Journal* **62**(2), 382–393.

Kim, H. E., Morse, A. and Zingales, L. (2009), 'Are Elite Universities Losing their Competitive Edge ?', *Journal of Financial Economics* **93**(3), 353–381.

Laband, D. N. and Tollison, R. D. (2000), 'Intellectual Collaboration', *Journal of Political Economy* **108**(3), 632–662.

Lazear, E. P. and Rosen, S. (1990), 'Male-Female Wage Differentials in Job Ladders', *Journal of Labor Economics* **8**(1), 106–123.

Leamer, E. (2007), 'A Flat World, a Level Playing Field, a Small World After All, or None of the Above ? Review of Thomas L. Friedman's *The World is Flat*', *Journal of Economic Literature* **45**(1), 83–126.

Lotka, A. (1926), 'The Frequency Distribution of Scientific Productivity', *Journal of the Whashington Academy of Sciences* **16**(12), 317–23.

Lovell, M. C. (1973), 'The Production of Economic Literature : An Interpretation', *Journal of Economic Literature* **11**(1), 27–55.

Lubrano, M., Bauwens, L., Kirman, A. and Protopopescu, C. (2003), 'Ranking European Economic Departments : A Statistical Approach', *Journal of the European Economic Association* **1**(6), 1367–1401.

MacKinnon, J. G., White, H. and Davidson, R. (1983), 'Tests for Model Specification in the Presence of Alternative Hypotheses', *Journal of Econometrics* **21**(1), 53–70.

Manning, W. G. and Mullahy, J. (2001), 'Estimating Log Models : To Transform or Not to Transform ?', *Journal of Health Economics* **20**(4), 461–494.

Mansfield, E. (1995), 'Academic Research Underlying Industrial Innovations : Sources, Characteristics, and Financing', *Review of Economics and Statistics* **77**(1), 55–65.

Mansfield, E. (1998), 'Academic Research and Industrial Innovation : An Update of Empirical Findings', *Research Policy* **26**(7-8), 773–776.

Manski, C. (1993), 'Identification of Endogenous Social Effects : The Reflection Problem', *Review of Economic Studies* **60**(3), 531–542.

Marshall, A. (1890), *Principles of Economics*, Macmillan, London.

Martin, W. and Pham, C. S. (2008), 'Estimating the Gravity Model when Zero Trade Flows are Frequent', *Mimeo* .

Martìnez-Zarzoso, I., Nowak-Lehmann D., F. and Vollmer, S. (2009), 'The Log of Gravity Revisited', *Mimeo* .

McDowell, J. M., Singell, L. D. and Ziliak, J. P. (2001), 'Gender and Promotion in the Economics Profession', *Industrial and Labor Relations Review* **54**(2), 224–244.

McDowell, J. M. and Smith, J. K. (1992), 'The Effect of Gender-sorting on Propensity to Coauthor : Implication for Academic Promotion', *Economic Inquiry* **30**(1), 68–82.

Medoff, M. H. (2003), 'Collaboration and the Quality of Economics Research', *Labour Economics* **10**(5), 597–608.

Melo, P. C., Graham, D. J. and Noland, R. B. (2009), 'A Meta-analysis of Estimates of Urban Agglomeration Economies', *Regional Science and Urban Economics* **39**(3), 332–342.

Moretti, E. (2011), Local labor markets, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 4B, North-Holland, Amsterdam, pp. 1237–1313.

Moulton, B. R. (1990), 'An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units', *Review of Economics and Statistics* **72**(2), 334–338.

Oaxaca, R. (1973), 'Male-Female Wage Differentials in Urban Labor Markets', *International Economic Review* **14**(3), 693–709.

Oyer, P. (2006), 'Initial Labor Market Conditions and Long-Term Outcomes for Economists', *Journal of Economic Perspectives* **20**(3), 143–160.

Petrongolo, B. and Olivetti, C. (2008), 'Unequal Pay or Unequal Employment ? A Cross-Country Analysis of Gender Gaps', *Journal of Labor Economics* **26**(4), 621–654.

Rosenblat, T. S. and Mobius, M. M. (2004), 'Getting Closer or Drifting Apart ?', *Quarterly Journal of Economics* **119**(3), 971–1009.

Rosenthal, S. S. and Strange, W. C. (2004), Evidence on the Nature and Sources of Agglomeration Economies, *in* V. Henderson and J.-F. Thisse, eds, 'Handbook of Regional and Urban Economics', Vol. 4, North-Holland, Amsterdam, pp. 2119–2171.

Sabatier, M. (2010), 'Do Female Researchers Face a Glass Ceiling in France ? A Hazard Model of Promotions', *Applied Economics* **42**(16), 2053–2062.

Sacerdote, B. (2011), Peer Effects in Education : How Might They Work, How Big Are They and How Much Do We Know Thus Far ?, *in* S. M. Eric A. Hanushek and L. Woessmann, eds, 'Handbook of the Economics of Education', Vol. 3, Elsevier, Amsterdam, pp. 249 – 277.

Santos Silva, J. and Tenreyro, S. (2006), 'The Log of Gravity', *The Review of Economics and Statistics* **88**(4), 641–658.

Sauer, R. D. (1988), 'Estimates of the Returns to Quality and Coauthorship in Economic Academia', *Journal of Political Economy* **96**(4), 855–866.

Schreiber, M. (2008), 'To Share the Fame in a Fair Way, HM Modifies H for Multi-authored Manuscripts', *New Journal of Physics* **10**, 040201–1–8.

Stigler, G. J. and Friedland, C. (1975), 'The Citation Practices of Doctorates in Economics', *Journal of Political Economy* **83**(3), 477–507.

Waldinger, F. (2012), 'Peer Effects in Science - Evidence from the Dismissal of Scientists in Nazi Germany', *Review of Economic Studies* **79**(2), 838–861.

Westerlund, J. and Wilhelmsson, F. (2011), 'Estimating the Gravity Model Without Gravity Using Panel Data', *Applied Economics* **43**(4/6), 641–649.

Winter-Ebmer, R. and Zweimuller, J. (1997), 'Unequal Assignments and Unequal Promotion in Job Ladders', *Journal of Labor Economics* **15**(1), 43–71.

Wooldridge, J. M. (1999), Quasi-Likelihood Methods for Count Data, *in* M. H. Pesaran and P. Schmidt, eds, 'Handbook of Applied Econometrics Volume II : Microeconomics', Oxford : Blackwell, pp. 352–406.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA. : The MIT Press.

Xiong, B. and Chen, S. (2012), 'Estimating Gravity Equation Models in the Presence of Heteroskedasticity and Frequent Zeros', *Mimeo* .

Zinovyeyav, N. and Bagues, M. (2012), 'It's Not What You Know, But Who You Know : The Role of Connections in Academic Promotions'. mimeo.