



HAL
open science

Highly-skilled migration : knowledge spillovers and regional performances

Claudia Noumedem Temgoua

► **To cite this version:**

Claudia Noumedem Temgoua. Highly-skilled migration : knowledge spillovers and regional performances. Economics and Finance. Université de Bordeaux, 2018. English. NNT : 2018BORD0173 . tel-02066507

HAL Id: tel-02066507

<https://theses.hal.science/tel-02066507>

Submitted on 13 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESIS PRESENTED FOR THE

PHD PROGRAM:

UNIVERSITE DE BORDEAUX

ECOLE DOCTORALE ENTREPRISE, ECONOMIE, SOCIETE

Spécialité : Sciences économiques

Candidate: Claudia NOUMEDEM TEMGOUA

**HIGHLY-SKILLED MIGRATION: KNOWLEDGE
SPILLOVERS AND REGIONAL PERFORMANCES
FOUR ESSAYS**

Thesis supervisor: Francesco LISSONI

Co-supervisor: Stefano BRESCHI

Soutenu le 02 Octobre 2018

Committee members:

Rosina MORENO, Professeur des Universités, Université de Barcelone, Présidente du jury

Alessandra FAGGIAN, Professeur des Universités, Institut des Sciences de Gran Sasso, Rapporteur

Ernest MIGUELEZ, Chargé de recherche, CNRS, Université de Bordeaux, Examineur

Diego USECHE, Maître de conférences, Université de Rennes I, Examineur

Francesco LISSONI, Professeur des Universités, Université de Bordeaux, Directeur de thèse

Stefano BRESCHI, Professeur des Universités, Université Bocconi, Co-Directeur de thèse

Titre : Migration hautement qualifiée: externalités de connaissances et performances régionales

Résumé : Les travailleurs hautement qualifiés - scientifiques et ingénieurs en particulier - en tant que acteurs majeurs de la production de connaissances et de l'innovation représentent un atout considérable pour la croissance économique. C'est ce qui justifie d'une part les efforts compétitifs déployés par plusieurs pays afin de les attirer et d'autre part la perception d'une perte en capital humain pour les pays d'origine. Cependant, ce phénomène est beaucoup plus complexe qu'il n'y paraît, car ces flux migratoires ont le potentiel de créer des retombées positives vers les pays d'origine ainsi que d'intensifier les échanges des connaissances entre pays de destination. La présente dissertation explore ce potentiel en se focalisant sur trois thématiques. Premièrement, elle participe au débat grandissant sur l'influence des réseaux de migrants hautement qualifiés en termes de flux internationaux de connaissances vers les pays d'origine en testant l'hypothèse de l'existence d'une relation positive entre migrations d'inventeurs et flux de connaissances (mesurés par les citations de brevets) sur la période 1990-2010. Ensuite, elle examine le rôle des diasporas constituées de personnes hautement qualifiées d'origine chinoise et indienne dans l'internationalisation des réseaux de connaissances, pour un échantillon de pays de destination membres de l'OCDE. On montre que les pays de destination qui accueillent ces diasporas ont tendance à collaborer davantage à la production de publications scientifiques et de brevets. Les diasporas hautement qualifiées d'autres pays (tels que le Vietnam, le Pakistan et l'Iran) génèrent des effets similaires. Enfin, en exploitant une riche base de données de l'Agence nationale sur la recherche de l'Afrique du Sud (NFR), on trouve que les scientifiques migrants rapatriés en Afrique du Sud sont plus susceptibles que les non-migrants de citer des publications plus récentes dans leurs publications ; ce qui suggère que ces migrants de retour jouent un rôle-clé dans le rapprochement de leur pays à la frontière scientifique.

Mots clés: Migrants, Flux de connaissances, Innovation, Collaborations scientifiques et techniques, Inventeurs, Chercheurs.

Laboratoire de recherche

Groupe de Recherche en Economie Théorique et Appliquée GREThA-UMR CNRS 5113,
Avenue Léon Duguit, 33608 Pessac – FRANCE

Title: Highly-skilled migration: knowledge spillovers and regional performances

Abstract: Highly-skilled are an important asset for a nation insofar that they enter in the production of knowledge as a highly valuable resource for economic growth. That is why many countries have been competing to attract them. Receiving countries are in general pictured as the biggest winners from highly-skilled migration, while the latter is perceived as a loss for sending countries, developing countries in particular. However, this phenomenon might not be as simplistic as it seems to the extent that these migration flows have the potential to generate some benefits to the sending countries while spurring knowledge exchanges among destination countries. This dissertation explores this potential, by addressing three research questions. Firstly, it participates to the growing debate on the effects of highly-skilled migrant networks in terms of international knowledge flows to the sending countries; testing the hypothesis of a positive relationship between inventors' migration and knowledge flows (as measured by patent citations) for the period 1990-2010. Secondly, it investigates the role of highly-skilled diasporas from Chinese and Indian origin in the internationalization of knowledge networks, for a sample of OECD destination countries. It shows that OECD countries hosting Chinese and Indian diasporas tend to collaborate more on publications and patents production. Similar results are found for other highly-skilled diasporas (such as those from Vietnam, Pakistan and Iran). Lastly, it exploits a rich database from the South African National Research Foundation (NRF) for a study on the brain gain potential of scientists' return migration. It is found that South African scientists with past-migration experience are more likely to cite more recent literature in their publications upon their return than non-migrant ones. This suggests these returnees play a key role in driving their country towards the knowledge frontier.

Keywords: Migrants, Knowledge flows, Innovation, Science and technology collaboration, Inventors, Scientists

Research Unit

Research Group Theoretical and Applied Economics, GREThA-UMR CNRS 5113,
Avenue Léon Duguit, 33608 Pessac – FRANCE

A tous ceux qui sont partis,
Portés par les eaux et le sable
On ne vous verra jamais plus ici
Mais vous êtes enfin libérés
Libérés de l'hypocrisie de ce monde
Monde injuste qui construit des barrières
Que cette injustice ne vous empêche de partir en paix
Et de grâce ne vous gênez surtout pas de souffler très fort sur ces barrières.

&

A mes parents

Acknowledgements

Completing this dissertation wouldn't have been possible without the support of several individuals.

First, I would like to thank Professor Francesco Lissoni who has given me the opportunity to start this incredible adventure. He has provided with much guidance, sometimes beyond what his role as my supervisor required. He has carried out this role with enough patience, understanding and presence, all what was necessary for me to complete this work and to make this experience as interesting as it has been.

I would like to thank all members of my thesis committee, Stefano Breschi, Alessandra Faggian, Rosina Morena, Ernest Miguelez, Diego Useche, for their support and the time they spend assessing this dissertation.

I would like to thank the GREThA for providing me with adequate resources and work environment. My gratitude goes to all the admirable colleagues I have had the opportunity to exchange and work with, particularly Ernest for his great support, Valerio, Erdem and Edoardo for their wonderful advices and friendship.

I would like to thank the IdEx Bordeaux programme for funding this work and easing its international visibility. Through its promotion of international doctorates, I have benefited from a jointly supervised thesis with Bocconi University.

I would be eternally grateful to my parents Noumedem and Donfack for everything. I want to thank Elodie for always being there; thank you to Junior, Vanessa, Maurelle and Marilyn for reminding me who I am.

Extended summary in French - Résumé étendu en Français :

Selon un récent rapport de la Banque Mondiale sur les migrations, en 2010 50% des migrants en âge de travailler dans 27 pays membres de l'Organisation de coopération et de développement économiques (OCDE) étaient diplômés du supérieur, contre 27% en 1990 (World Bank, 2018). Un fait majeur ressort de ces chiffres: les flux migratoires sont de plus en plus constitués de personnes hautement qualifiées. Ceci est principalement dû à une offre accrue de travailleurs hautement qualifiés à l'échelle mondiale, ainsi qu'à une demande croissante pour ces derniers, venant en particulier des pays développés. Cependant, ces faits contrastent avec les croyances et idées diffusées par les vagues de propagande anti-migration au cœur des débats politiques et alimentés par les médias traditionnels en Europe et aux États-Unis; et qui gagne progressivement de l'ampleur à travers le monde. Cette tendance à se focaliser sur le côté défavorable des migrations n'est pas un fait nouveau et est loin d'être l'apanage des politiciens et des médias. En effet, une approche similaire a été adoptée par diverses théories et travaux empiriques dans différentes disciplines scientifiques parmi lesquelles l'Economie. Il existe dans cette discipline scientifique un courant d'idées bien établi dans la littérature sur les migrations qui soulève diverses problématiques ciblées sur les migrations des pays en développement vers les pays développés. Ces problématiques sont entre autres: l'hypothèse selon laquelle les migrants substitueraient les nationaux sur le marché du travail dans les pays de destination, encore appelé effet d'éviction ou encore qu'ils entraîneraient la baisse des salaires (Borjas, 2003, 2004); et l'éventuelle perte en capital humain pour les pays d'origine, également connue sous le nom de fuite des cerveaux (Bhagwati & Hamada, 1974; Grubel & Scott, 1966). Cette dernière problématique tient de l'ordre des idées reçues dans la littérature sur les migrations des travailleurs hautement qualifiés, notamment des pays les moins développés vers les pays développés; ceci compte tenu du taux de perte relativement élevé de ces travailleurs dans ces pays (Lowell & Findlay, 2001).

Cependant, ce courant d'idées a tendance à omettre les retombées significatives sous la forme de mécanismes de compensation découlant des migrations hautement qualifiées. Les potentiels gains nets bénéficieraient non seulement aux pays d'origine, mais aussi aux pays d'accueil. Ceci constitue l'argument principal à la base de la théorie du brain gain, une

nouvelle ligne de recherche préconisée par les récents courants sur les migrations (Stark et al., 1997, 1998; Stark & Wang, 2002). Cette littérature identifie plusieurs facteurs déterminants de brain gain découlant des migrations des personnes hautement qualifiées. Des études antérieures ont mis en exergue le fait que les perspectives de migration inciteraient à accroître les investissements dans l'éducation et la formation du capital humain dans les pays d'origine (Beine et al., 2001; Massey et al., 1993; Stark & Wang, 2002); D'autres ont montré l'importance des flux de fonds, de capitaux et de produits par la diaspora des pays d'origine (Ascencio, 1993; Massey & Parrado, 1994; Taylor, 1999). De plus, avec les flux croissants des migrations de personnes hautement qualifiées une autre forme de flux comme instrument essentiel des brain gain a progressivement émergé, à savoir les retours de connaissances. Ces derniers sont perçus comme la contribution des migrants à la production des connaissances et à l'innovation dans les pays d'origine. Les retours de connaissances peuvent prendre trois principales formes: les retombées de connaissances provenant des liens des migrants hautement qualifiés avec leurs pays d'origine; la contribution directe des migrants rapatriés vers leurs pays d'origine; et les réseaux des diasporas des différents groupes ethniques à l'intérieur des pays d'accueil (Breschi et al., 2016). Une caractéristique commune à ces différentes formes demeure l'importance des liens sociaux, autant sur le plan professionnel qu'ethnique. Ceci s'explique par le fait que l'échange de connaissances tacites requiert une certaine proximité physique ou un contact direct (Breschi & Lissoni, 2009; Jaffe et al., 1993; Miguelez, 2016). Les liens sociaux reposent sur des mécanismes complexes de proximité culturelle, linguistique et sociale produisant des réseaux de migrants susceptibles de déboucher sur des collaborations et échanges scientifiques et technologiques entre pays d'origine et pays de destination, mais aussi à l'intérieur des pays de destination (Breschi et al., 2017; Jonkers & Tijssen, 2008; Scellato et al., 2015).

Des études sur le brain gain ont mis en évidence la surreprésentation des migrants hautement qualifiés dans les domaines innovants dans leurs pays d'accueil (Chellaraj et al., 2008; Stephan & Levin, 2001); ainsi que la capacité des réseaux de migrants à favoriser la diffusion des connaissances à l'intérieur de ces pays (Agrawal et al., 2008; Kerr, 2009). Cependant, ces effets bénéficieraient uniquement à un groupe restreint de pays de destination, dans la mesure où les migrants hautement qualifiés sont concentrés dans ces pays; tous membres de l'OCDE. Cette dernière englobe les pays où les activités liées à l'innovation et à la production de connaissances sont les plus intenses (De Backer & Basri, 2008); et à l'exception des États-Unis, du Canada et de l'Australie, ces pays sont majoritairement européens. Cependant, dans

la littérature sur les migrations de personnes hautement qualifiées et l'innovation les productions scientifiques sur les pays européens sont rares, car ce champ de recherche est largement dominé par des études sur les États-Unis en tant que pays de destination, avec l'Inde et la Chine comme pays d'origine. De plus, le manque d'intérêt pour les pays africains représente un frein majeur à l'avancée des travaux de recherche dans ce domaine, compte tenu du fait que ces pays sont parmi les pays les plus touchés par la fuite des cerveaux (Docquier & Rapoport, 2012). Tout ceci souligne l'existence de multiples zones d'ombres sur lesquelles il serait nécessaire de se pencher véritablement dans l'optique de mieux assimiler les mécanismes autour de cette thématique. Cela requiert initialement l'évaluation des différentes catégories de migrants hautement qualifiés et de leurs caractéristiques respectives.

Les migrants hautement qualifiés appartiennent à diverses catégories dont les plus communément étudiées dans la littérature sur les migrations et l'innovation sont les scientifiques et les ingénieurs ; en raison de leur position centrale dans la création de connaissances et l'innovation. Ensuite viennent les étudiants en doctorat – du fait d'une meilleure documentation de cette catégorie. Les études sur les migrations des travailleurs hautement qualifiés et leur contribution à la diffusion des connaissances et à l'innovation ont produit des résultats assez mitigés pour chacune de ces catégories. Ces résultats varient en fonction du cadre de la recherche et du type de données utilisées. De plus, la qualité et la disponibilité des données sont des facteurs décisifs de l'intérêt porté aux catégories ci-dessus mentionnées. Les études empiriques à grande échelle sont rares et reposent en grande partie sur les données sur les brevets (Almeida et al., 2010; Foley & Kerr, 2013; Kerr, 2007, 2008, 2009). Cependant, il existe des études à échelle plus réduite exploitant d'autres sources de données ; notamment des études basées sur des données de publications (Baruffaldi & Landoni, 2012; Jonkers & Cruz-Castro, 2013; Scellato et al., 2015); ou encore des productions scientifiques s'appuyant sur des données sur l'immigration de doctorants/post-docs afin d'évaluer leurs performances (Hunt, 2009; Hunt & Gauthier-Loiselle, 2010).

Dans la présente dissertation une approche extensive et globale est adoptée, réunissant une couverture diversifiée de différentes catégories de migrants hautement qualifiés ; ainsi que des sources de données avec une large couverture de pays d'origine et d'accueil. Ces données sont traitées selon différentes approches méthodologiques applicables à chaque thématique abordée. Ces dernières relèvent de points critiques à ce jour inabordés dans la littérature – ou encore qui demandent une réflexion plus approfondie -, notamment du fait de contextes géographiques ou régionaux particuliers. Sans toutefois faire fi des pièges potentiels des

migrations de masse des travailleurs hautement qualifiés, le choix de l'orientation du débat est plutôt porté sur la question des potentiels gains pouvant résulter des migrations de travailleurs hautement qualifiés. Les données sur ces flux migratoires montrent que la dynamique de la mobilité des travailleurs hautement qualifiés semble imparable (Barro & Lee, 2013; Docquier et al., 2009). Par conséquent, l'accent devrait être mis sur la formulation d'une meilleure stratégie pour mieux appréhender cette dynamique, mais aussi l'origine et la nature des gains. Ceci permettra d'aider à la formulation de recommandations politiques adéquates. Suivant cette logique, la présente thèse s'organise en quatre chapitres.

Le premier chapitre¹ consiste en une revue de la littérature sur la relation entre migrations et l'innovation ; avec une discussion poussée autour des principales études sur l'impact des migrations des travailleurs hautement qualifiés sur les pays de destination et d'origine.

Le deuxième chapitre² participe au débat grandissant sur l'influence des réseaux de migrants hautement qualifiés en termes de flux internationaux de connaissances vers les pays d'origine aussi bien que vers les pays d'accueil. Dans un premier temps, une analyse des retours de flux de connaissances vers les pays d'origine provenant des inventeurs immigrés – une catégorie représentative des migrants hautement qualifiés composée en grande partie de scientifiques et ingénieurs – est effectuée. Ensuite vient l'analyse des flux de connaissances vers les pays d'accueil découlant de l'immigration d'inventeurs. Un modèle de gravité est utilisé afin de tester l'hypothèse de l'existence d'une relation positive entre flux de connaissances et migrations d'inventeurs sur la période 1990-2010. Les citations de brevets sont utilisées comme proxy des flux internationaux de connaissances. Les résultats obtenus confirment l'hypothèse initiale. Plus précisément, ces résultats montrent qu'en doublant le nombre d'inventeurs d'une certaine origine à l'intérieur d'un pays d'accueil les flux de connaissances vers leur pays d'origine observeront une hausse de 8,3% tandis que la hausse de flux de connaissances vers ce pays de d'accueil sera de 6%.

Le troisième chapitre aborde la question du rôle des diasporas constituées de personnes hautement qualifiées d'origine chinoise et indienne dans l'internationalisation des réseaux de connaissances, pour un échantillon de pays de destination membres de l'OCDE. Plus précisément, deux principaux types de réseaux de connaissances sont analysés: les réseaux de co-inventeurs et de co-auteurs. Des données à l'échelle nationale sur les migrants hautement qualifiés tirées de la base de données OCDE-DIOC (Base de données sur les immigrés dans

¹ Déjà publié en tant que chapitre d'un livre, et avec déjà comme co-auteurs Stefano Breschi et Francesco Lissoni.

² Co-écrit avec Ernest Miguelez.

les pays de l'OCDE, 2000/01, 2010/11) et des informations sur les réseaux de co-auteurs et de co-inventeurs provenant des publications et des brevets sont utilisées conjointement. L'analyse de ce chapitre est basée sur des régressions de modèle de gravité. Les résultats obtenus à l'issue de ces régressions montrent que les pays de destination avec une part importante des diasporas indiennes ou chinoises de personnes hautement qualifiées ont tendance à collaborer davantage sur les publications et les brevets. En étendant l'analyse à d'autres pays, des résultats similaires sont obtenus pour le cas du Vietnam, du Pakistan et de l'Iran.

Dans le dernier chapitre³, le potentiel en brain gain des scientifiques migrants qui se sont rapatriés dans leur pays d'origine est examiné. Ce chapitre exploite une riche base de données de l'Agence nationale sur la recherche de l'Afrique du Sud (NFR), regroupant des informations personnelles et professionnelles sur les scientifiques locaux. Ces données sont connectées aux informations bibliométriques de la base de données Web of Science. A l'issue d'une analyse reposant sur un modèle à effets aléatoires, les résultats obtenus montrent que les scientifiques migrants rapatriés en Afrique du Sud sont plus susceptibles que les non-migrants de citer des publications plus récentes dans leurs publications. Des tests de robustesse, notamment ceux liés aux problèmes d'endogénéité, sont effectués au moyen d'une analyse basée sur la méthode des différences de différences appliquée à un échantillon réduit.

Dans l'ensemble, la présente dissertation confirme l'effet de brain gain en termes de diffusion des connaissances et la contribution à l'innovation des migrants hautement qualifiés, tant dans les pays de destination que dans les pays d'origine. Elle met un accent particulier sur le rôle crucial des interactions sociales au sein des réseaux ethniques et professionnels, de différents groupes et catégories de migrants hautement qualifiés. Les deuxième et troisième chapitres sont pertinents d'un point de vue politique dans la mesure où ils soulignent l'importance du maintien et du renforcement des liens entre membres de diaspora, aussi bien entre pays de destination qu'entre pays d'origine et de destination. Quant au dernier chapitre, il met en exergue l'importance pour les chercheurs d'acquérir une certaine expérience à l'étranger, ce qui participe à accroître leurs performances; mais aussi représente un élément déterminant de leur intégration dans les réseaux de connaissances transnationaux - et indirectement de l'intégration des institutions de leurs pays d'origine.

En ce qui concerne les futurs plans, il est prévu une amélioration du chapitre 4 qui nécessite un travail supplémentaire de traitement des données, notamment la connexion des données

³ Co-écrit avec Robin Cowan, Moritz Muller et Francesco Lissoni.

NFR à de plus amples informations bibliométriques. Ensuite, il serait intéressant d'évaluer d'autres indicateurs des performances des chercheurs sud-africains.

Contents

- List of Figures ----- 14**
- List of Tables ----- 15**
- 1 Migration and innovation: a survey of recent studies ----- 22**
 - 1.1 Introduction ----- 22
 - 1.2 Highly-skilled migration: general evidence ----- 24
 - 1.3 Migration and innovation in destination countries ----- 26
 - 1.4 Migration and innovation in origin countries: “Knowledge remittances” from hs migrants ----- 29
 - 1.5 International mobility of hs workers within companies and innovation ----- 33
 - 1.6 Discussion and conclusions ----- 37
- 2 Immigration externalities, knowledge flows and brain gain ----- 39**
 - 2.1 Introduction ----- 39
 - 2.2 Background literature ----- 40
 - 2.3 Methodology and data ----- 45
 - 2.3.1 Empirical approach ----- 45
 - 2.3.2 Data ----- 46
 - 2.3.2.1 Patent citations as a measure of knowledge flows ----- 46
 - 2.3.2.2 The international mobility of inventors ----- 48
 - 2.3.2.3 Control variables ----- 49
 - 2.3.2.4 Descriptive statistics ----- 51
 - Citations and migration corridors ----- 51
 - Trends of citations and inventors’ migration in selected countries* ----- 54
- 2.4 Results ----- 58
 - 2.4.1 Skilled migration and brain gain ----- 58
 - 2.4.1.1 Baseline estimations ----- 58
 - 2.4.1.2 Are the biggest players driving our results? ----- 61
 - 2.4.1.3 Are there differences between income group pairs? ----- 63
 - 2.4.1.4 Instrumental variable strategy ----- 66
 - 2.4.2 Immigration and innovation ----- 69
 - 2.4.2.1 Baseline estimations for knowledge inflows ----- 69
 - 2.4.2.2 Are the biggest players driving our results? ----- 70
 - 2.4.2.3 Are there any differences between income group pairs? ----- 72
 - 2.4.2.4 Instrumental variable strategy for the inflows of knowledge ----- 75

2.5 Discussion and conclusions -----	77
Appendix A – Some additional tables-----	78
3 Highly-skilled migration and the internationalization of knowledge-----	85
3.1 Introduction -----	85
3.2 Literature review-----	86
3.2.1 Migration, social networks and innovation -----	86
3.2.2 Science and technology (S&T) collaborations -----	90
3.3 Key definitions -----	92
3.4 Methodology-----	94
3.4.1 Empirical approach-----	94
3.4.2 Data -----	96
3.4.2.1 Dependent variables -----	96
3.4.2.2 Explanatory variables and controls-----	98
3.4.3 Descriptive statistics-----	102
3.5 Results -----	106
3.5.1 The effect of the Chinese and Indian hs diaspora in S&T collaborations -----	106
3.5.2 Some robustness check-----	113
3.6 Concluding remarks -----	120
Appendix B – Some tables -----	122
Appendix C – IV strategy-----	129
Appendix D – Exploring R&D cooperation-----	131
4 Return migration and citations recency: the case of South African researchers ----	139
4.1 Introduction -----	139
4.2 Literature review-----	142
4.3 Methodology and data -----	145
4.3.1 Specifications-----	145
4.3.2 Data-----	147
4.3.2.1 Publications-----	147
4.3.2.2 Migration pattern -----	150
4.3.2.3 Returnees’ vs. non-migrants’ characteristics -----	152
4.4 Results -----	154
4.4.1 Baseline results: panel data analysis -----	154
4.4.2 Robustness check: conditional dif-in-dif analysis-----	162
4.4.2.1 Propensity score matching-----	163
4.4.2.2 CDID regression results -----	166
4.4.3 Migration effects over time-----	170
4.4.3.1 Delay/decay in the post-migration effect-----	170

4.4.3.2 First and subsequent migration events differentiated-----	177
4.5 Concluding remarks -----	180
Appendix E – Some tables-----	182
Appendix F – Discussion on the conceptualization of returnees-----	184
Appendix G – Normalizing recency over scientific field average-----	188
Bibliography-----	189

List of Figures

Figure 2.1: Migration and knowledge trends comparison for China and the US	56
Figure 2.2: Migration and knowledge trends comparison for India and the US	57
Figure 3.1: NOHS diasporas Vs. OHS migrant bilateral links.....	94
Figure 3.2: Chinese hs migrants distribution in 31-OECD destination countries: DIOC 10/11 vs. DIOC 00/01	105
Figure 3.3: Indian hs migrants distribution in DIOC 10/11 and DIOC 00/01	106
Figure 3.4: Co-inventorship elasticities of different NOHS diasporas	117
Figure 3.5: Co-authorship elasticities of different NOHS diasporas	119
Figure D1: R&D cooperation elasticities of different NOHS diasporas.....	135
Figure 4.1: Trend in the average age of citations 1985-2014.....	156
Figure 4.2: Trend in the average recent citations ratio in % 1985-2014	156
Figure 4.3: Comparison of kernel distribution of the propensity scores.	165

List of Tables

Table 2.1: Sample of citations at the inventor level	47
Table 2.2: Knowledge flows variable.....	47
Table 2.3: Variables definition	51
Table 2.4: Citations corridors (Total flow of citations for the top 20 country pairs for the period 2006-2010)	53
Table 2.5: Migration corridors (Total of inventor immigrants for the top 20 country pairs for the period 2006-2010)	54
Table 2.6: Knowledge outflows PPML baseline estimations for the period 1990 – 2010	60
Table 2.7: Knowledge outflows PPML estimations without some of the biggest players for the period 1990 – 2010.....	62
Table 2.8: Knowledge outflows income group PPML estimations for the period	65
Table 2.9: First stage OLS: determinants of inventors’ migration	67
Table 2.10: IV poisson regressions for knowledge outflows	68
Table 2.11: Knowledge inflows from migrant inventors 1990 – 2010.....	70
Table 2.12: Knowledge inflows PPML estimations without some of the biggest players for the period 1990 – 2010	71
Table 2.13: Knowledge inflows income group PPML estimations for the period 1990 - 2010.....	74
Table 2.14: IV poisson regressions for knowledge inflows	76
Table A1: Citations corridors – without high income as citing countries (Total flow of citations for the top 20 country pairs for the period 2006-2010)	78
Table A2: Citations corridors – with low income only as citing countries (Total flow of citations for the top 20 country pairs for the period 2006-2010)	79
Table A3: Migration corridors without high income origin countries (Total of inventor immigrants for the top 20 country pairs for the period 2006-2010).....	80
Table A4: Migration corridors – with low income only as origin (Total of inventor immigrants for the top 20 country pairs for the period 2006-2010)	81
Table A5: Descriptive statistics.....	82
Table A6: PPML regressions with explanatory variables time lagged for the knowledge outflows.....	83
Table A7: More explanatory variables for the knowledge outflows model	84
Table 3.1a: Example of co-inventorship data (two patents).....	97
Table 3.1b: Count of co-inventorships, by country pairs from data in Table 3.1a	97
Table 3.2: Variables definition	101
Table 3.3: Science and technology research corridors for years 2010 – 2014	102
Table 3.4: Top fifteen receiving countries of Chinese hs migrants in DIOC 10/11 and 00/01	103
Table 3.5: Top fifteen receiving countries of Indian hs migrants in DIOC 10/11 and 00/01.....	104
Table 3.6: Chinese and Indian hs diaspora in international co-inventorship.....	109
Table 3.7: Chinese and Indian hs diaspora in international co-authorship.....	112
Table 3.8: Other Chinese and Indian skill groups and international Co-inventorship and Co-authorship	115
Table B1: Detailed migration data sources by country of destination.....	122

Table B2: Descriptive statistics	123
Table B3: Marginal effects of co-inventorship and co-authorship	123
Table B4: Other NOHS diaspora co-inventorship elasticities	124
Table B5: Other NOHS diaspora co-authorship elasticities	126
Table B6: Emigrant population 15+ in the OECD in 2010/11 by country origin	128
Table C1: GMM estimates with instrumented Chinese and Indian hs diasporas	130
Table D1: Counting of country pairs R&D cooperation	132
Table D2: Sample of R&D projects under the FP5	132
Table D3: Chinese and Indian hs diaspora in international R&D cooperation	134
Table D4: Other NOHS diaspora impact on R&D cooperation.....	137
Table 4.1: RSA returnees 1985-2014: PhD, Master degree and work experience	152
Table 4.2: Sample characteristics	153
Table 4.3: Top host countries between 1985-2014.....	154
Table 4.4: Sample publications with the average age of citations.....	155
Table 4.5: Variables definition	158
Table 4.6: Return migration and citations recency	160
Table 4.7: Descriptive statistics for the unmatched and matched samples.....	164
Table 4.8: Mean and Median standardized bias for the matched and unmatched samples	164
Table 4.9: Summary statistics of RSA migrant researchers and their control.....	166
Table 4.10: Results from the difference-in-difference regression	168
Table 4.11: Mean output differences in the treatment and control groups	169
Table 4.12: Migration/return events distribution.....	170
Table 4.13: Summary statistics of migration/return	170
Table 4.14: Results from regressions with post migration effect delay	173
Table 4.15: Results from regressions with post migration effect decay	175
Table 4.16: Distribution of the number of migration/return events by the duration of stay abroad....	177
Table 4.17: Results from regressions with post migration dummy first and second trips.....	179
Table E1: Descriptive statistics	182
Table E2: Correlation table	183
Table E3: Logit results from the PSM analysis.....	183
Table F1: Baseline regressions without researchers with less than a year long migration experience	186
Table F2: Regressions with PhD experience abroad.....	187
Table G1: Results from regressions with recency normalized over scientific field average.....	188

Introduction

A recently released World Bank report on migration shows in 2010, 50 percent of the working-age migrants in 27 high-income countries of the Organization for Economic Cooperation and Development (OECD) were tertiary-educated as compared with 27 percent tertiary-educated migrants in 1990 (World Bank, 2018). These figures highlight one crucial element which can be resumed as follows: migration has become increasingly highly-skilled. This is mainly due to an increased supply of highly-skilled globally, as well as a growing demand for them particularly from developed economies. All this is however in contrast with beliefs and ideas diffused by the wave of anti-migration propaganda that has been at the core of political debates and fuelled by mainstream media in Europe and the US; and which is progressively spreading worldwide. This tendency to stress on the unfavourable side of migration is neither contemporary and nor solely common with politics and media. Indeed, various theories and empirical work from different disciplines in the scientific world have followed – and in some cases supported – it. In economics, there is a well-established trend of literature that have emphasized on migration from less developed to developed countries, raising several concerns. The most common ones are: the idea of migrants' substituting locals in the receiving economy labour market, the so-called crowding-out effect, or lowering wages (Borjas, 2003, 2004); and the potential loss of human capital for sending countries also known as brain drain (Bhagwati & Hamada, 1974; Grubel & Scott, 1966). The latter concern has been established as a received wisdom in the literature of highly-skilled migration particularly from less developed to developed countries insofar that the loss of highly-skilled is relatively high in the former group (Lowell & Findlay, 2001).

However, this literature tends to ignore the significant efficiency gains in the form of compensating mechanisms that might derive from highly-skilled migration. The resulting potential net gains apply not only to sending countries but also to the receiving ones. This constitutes the main argument behind the brain gain theory, a new line of research advocated by recent trend of literature on migration (Stark et al., 1997, 1998; Stark & Wang, 2002). This literature identifies several channels of brain gain from highly-skilled migration. Prior studies have stressed on higher incentives to invest on education and human capital formation in sending countries led by the prospect of migration (Beine et al., 2001; Massey et al., 1993; Stark & Wang, 2002); while others have pointed out to the importance of product and

financial remittances sent back home by home countries' diaspora (Ascencio, 1993; Massey & Parrado, 1994; Taylor, 1999). Besides with the increasing intensity of highly-skilled migration, another form of remittances has emerged as a critical instrument of brain gain, namely "knowledge remittances". The latter is intended as migrant's contribution to knowledge formation and innovation in origin countries. These knowledge remittances may come in three forms: knowledge spillovers from highly-skilled migrants' ties with their home country; returnees' direct contribution home; and diaspora networks formed by ethnic group members within host countries (Breschi et al., 2016). Yet one common feature to these different forms remains the key importance of work-related and ethnic-related social ties – as tacit knowledge exchange requires physical proximity or face to face contact (Breschi & Lissoni, 2009; Jaffe et al., 1993; Miguelez, 2016). These ties rely on complex mechanisms involving cultural, language and social proximity, which create migrant networks that may lead to scientific and technological collaboration or exchanges between home and host countries, but also within host countries (Breschi et al., 2017; Jonkers & Tijssen, 2008; Scellato et al., 2015).

Besides evidences on highly-skilled migrants over-representation in innovative activities in their host countries (Chellaraj et al., 2008; Stephan & Levin, 2001), migrant networks have been found to favour knowledge diffusion within host countries (Agrawal et al., 2008; Kerr, 2009). However, these effects might not evenly benefit to all destinations insofar that only a reduced group of countries hosts the highest share of highly-skilled migrants; all of them belonging to OECD. Admittedly, countries with the most intensive innovation and knowledge production activities are OECD members (De Backer & Basri, 2008); and except from the US, Canada and Australia the large majority of them is from Europe. Surprisingly, there are less studies on European countries as the highly-skilled migration and innovation literature has been dominated by the US as a destination country, with highly-skilled migrants from India and China. Moreover, the lack of focus on African countries represents a major drawback, since migration figures show these countries are among the ones which suffer the most from brain drain (Docquier & Rapoport, 2012). All the above highlights some critical shortages in this literature. Hence, central to the goal of deepening our understanding of all the mechanisms at work within this line of research lies the necessity to overcome these shortages. This starts with assessing the various categories of highly-skilled migrants and their specificities.

Highly-skilled migrants come from various backgrounds such as engineers, entrepreneurs, scientists, teachers, graduate students...etc (World Bank, 2018). But the most commonly investigated ones in the highly-skilled migration and innovation studies are scientists and engineers – partly due to their stand as leading actors in knowledge creation and innovation –; and also graduate students – a well documented highly-skilled category. There are mixed evidences coming from highly-skilled migration studies on each of these categories' contribution to knowledge diffusion and innovation, depending on the research scope and the type of data used. Admittedly, the focus on the aforementioned categories is also partly led by data quality and availability reasons. Large scale quantitative evidence is scant, and largely relies on patent data studies (Almeida et al., 2010; Foley & Kerr, 2013; Kerr, 2007, 2008, 2009). There are however some cases where other sources of smaller scale data are exploited, such as studies based on publications data (Baruffaldi & Landoni, 2012; Jonkers & Cruz-Castro, 2013; Scellato et al., 2015) or papers relying on PhD students/post-docs immigration records for assessing their performances as compared with natives (Hunt, 2009; Hunt & Gauthier-Loiselle, 2010).

In the present thesis, I adopt an integrative approach which involves basing my research on a rich coverage of highly-skilled migrant categories, multiple dataset with a wide range of home and host countries and relevant methodology. I address critical questions that were still unexplored, particularly within specific geographical or regional settings. Acknowledging all potential pitfalls of massive highly-skilled migration, I rather choose to draw the attention around possible benefits that may arise from highly-skilled migration. Migration patterns figures show the dynamics of highly-skilled mobility seems rather unstoppable (Barro & Lee, 2013; Docquier et al., 2009). Therefore I believe much effort should be made in order to find the best approach to understand this dynamics, the origin and nature of its gains; and to help in the formulation of adequate policy recommendations in this regard. The present dissertation is thus organized into four chapters.

The first one⁴ is a survey of literature on the relationship between migration and innovation, with a discussion of relevant studies on the impact of highly-skilled migration on destination and origin countries.

The second chapter⁵ adds to the growing literature on the influence of networks of highly-skilled migrants on international knowledge flows, in migrants' home as well as host

⁴ Already published as a chapter of a book and co-authored with Francesco Lissoni and Stefano Breschi.

countries. More precisely, it first explores knowledge feedbacks to home countries generated by migrant inventors, a representative category of highly-skilled migrants, most of them scientists and engineers. Second, it investigates the knowledge inflows to host countries brought by inventors. The hypothesis of a positive relationship between knowledge flows and highly-skilled migration is tested in a country-pair gravity model setting, for the period 1990-2010. I use cross-country citations to Patent Cooperation Treaty (PCT) patents – the patent database from the World International Patent Office (WIPO) – as a measure of international knowledge diffusion. Results confirm the initial assumption of a positive impact of highly-skilled migrants on knowledge flows to their homelands as well as to their host countries. I find that doubling the number of inventors of a given nationality at a destination country, leads to an 8.3% increase in knowledge outflows to their home economy from that same host land; while a similar increase in the number of migrant inventors produces a 6% increase in the knowledge inflows to the host economy.

The third chapter investigates the role of Chinese and Indian highly-skilled diaspora in the internationalization of knowledge networks, for a sample of OECD destination countries. Here I mainly focus on two types of knowledge networks: co-inventorship and co-authorship. I jointly exploit country-level data on highly-skilled migration from the OECD-DIOC database (Database on Immigrants in OECD Countries; 2000/01, 2010/11) and information on co-authorship and co-inventorship from publication and patent data. Based on a gravity model regression analysis, I find that OECD country pairs hosting sizeable portions of the Indian or Chinese highly-skilled diasporas tend to collaborate more on publications and patents production, after controlling for other migration trends. When extending the analysis to other countries, I find similar results for Vietnam, Pakistan and Iran.

In the last chapter⁶, I examine the brain gain potential of scientists' return migration for their home countries. I use a rich database from the South African National Research Foundation (NRF) recording local researchers' curricula, which is linked to bibliometric information from the Web of Science. Based on random effect panel data analysis, I find that, South African scientists with past-migration experience are more likely to cite more recent literature in their publications upon their return than non-migrant ones. Further robustness tests of the results with respect to endogeneity issues are conducted by means of a conditional difference-in-difference analysis of a case-control dataset.

⁵ Based on a paper co-authored with Ernest Miguelez.

⁶ Based on a paper co-authored with Robin Cowan, Moritz Muller and Francesco Lissoni.

All in all, this dissertation confirms the brain gain, knowledge diffusion and innovation contribution of highly-skilled migration, both in host and home countries. It has stressed on the critical role of social interactions within ethnic and professional networks for different migrant groups and highly-skilled categories. The second and third chapters are relevant from a policy perspective to the extent that they underline the importance of maintaining and strengthening ties among diaspora members not only within specific host countries but also across host-host and host-home countries. As for the last chapter, it points to the key position of researchers' foreign experience which participates in enhancing their performances; and so represents a critical determinant of their integration into transnational knowledge networks – and indirectly the integration of their home country institution upon their return.

Future plans involve further improvement of chapter 4 which requires some more data mining work, particularly with connecting the NFR data with bibliometric information. Then, it would be interesting to assess other indicators of performances from South African returnees.

Chapter 1

Migration and innovation: a survey of recent studies

1.1 Introduction

Migration and innovation are two phenomena whose ties date back a long time in history, well before the emergence of professional science and engineering (S&E). David (1993) historical excursus on the birth of modern intellectual property rights (IPRs) reminds us that the latter originate from the privileges granted by Italian states of the 14th-15th century to foreign craftsmen, in order to lure them away from their home countries (or rival states) and inject new techniques in the local industry. In the same years, Tudor England was engaging actively in «the negotiation ... of secret agreements designed to attract skilled foreign artisans into [the Crown's] service. German armourers, Italian shipwrights and glass-makers, French iron workers were enticed to cross the Channel in this fashion» (David, 1993). (Hornung, 2014) provides econometric evidence in favour of Frederich List's (1841) classic argument on the productivity impact of Huguenot migration from France to Prussia after the revocation of the Edict of Nantes in 1685. In a similar vein, Moser et al. (2014) show how Jewish professors of Chemistry seeking refuge from Nazi Germany were responsible for a significant growth in US patenting activity in chemical technologies, both directly and indirectly (by opening up new research avenues for US inventors).

What makes the study of migration and innovation a hot research topic nowadays is the steady increase in the global flows of scientists and engineers (S&Es) observed over the past 20 years, both in absolute terms and as a percentage of total migration flows (Docquier & Rapoport, 2012; Freeman, 2013). These flows have been fed by an increasing number of countries, most notably China, India, and Eastern Europe. This raises a number of questions on the role these migrants play in the innovation process in both their destination and origin countries.

The most common questions asked with reference to destination countries, most notably the United States, can be summarized as follows: Are foreign S&Es complements or substitutes

of local ones? In other words, do they increase their destination country's innovation potential, or do they simply displace the local workforce (Borjas, 2004; Chellaraj et al., 2008; Hunt & Gauthier-Loiselle, 2010)? Are destination countries increasingly dependent on the immigration of S&Es (including graduate students) to maintain their present technological leadership? Does such dependence require the implementation of dedicated immigration policies (Chaloff & Lemaitre, 2009)?

As for origin countries, the key research questions concern the extent of their loss of human capital ("brain drain") and the nature and effectiveness of potential compensating mechanisms, such as knowledge spillovers from destination countries or the contribution to local innovation by returnee S&Es and entrepreneurs (Agrawal et al., 2011; Kerr, 2008). In this respect, some debate exists on the role of intellectual property, most notably in the aftermath of many origin countries' subscription of TRIPs, the Trade Related Intellectual Property Agreements that come with the adherence to the World Trade Organization (Maskus & Fink, 2005).

While rich in questions, this emerging literature is still poor in answers. One important limitation concerns the empirical side, and the lack of extensive and detailed data for micro-econometric analysis. Another important limitation concerns its almost exclusive focus on one destination country, the US, and its top providers of foreign talents over the recent years, namely India, China, and other East Asian states. US-centrism is not peculiar to this field of studies, but here it bears the additional disadvantage of reducing a multi-polar phenomenon, one in which several countries act both as source and destination of migration flows, to a set of binary relationships between the US and a limited set of origin countries.

This chapter reviews existing quantitative studies that address, either directly or indirectly, the relationship between migration and innovation. We first consider general studies on the growing phenomenon of highly-skilled (tertiary educated) migration, as well as research on mobility of PhD graduate and scientists (section 2). We then move on to survey the existing evidence on migration's impact on destination countries, with special emphasis on specific categories of migrants, namely inventors and international science and technology students (section 3). Section 4 deals with the impact on origin countries: we review both some general evidence and selected region- or country-specific studies. In section 5 we review the special case of intra-company international mobility. Section 6 concludes.

1.2 Highly-skilled migration: general evidence

Studies on highly-skilled (hs) migration belong to a longstanding tradition of research on migration and development (for a survey of economic studies, see Docquier & Rapoport (2012); for a cross-disciplinary survey, see De Haas (2010)). The most comprehensive evidence on the phenomenon comes from the DIOC database family, which is based upon information for the 2000 census round, supplemented by labour force surveys, with updates for year 2005 and 2010 (related data for 1990 can be found in Artuç et al. (2015)). Hs migrants (and residents) are identified as those having a tertiary education degree, which is a diploma earned at college or university.⁷

DIOC data show that, at least since the 1990s, hs migration has grown both as a stock and as a share of total migration (Respectively: from around 13 million units worldwide in 1990 to 26 million units in 2005/06; and from 30% in 1990 to almost 40% in 2005/06). Besides, migration rates for the tertiary educated are higher than for the non-tertiary educated.

Origin countries of hs migrants are the largest ones, especially those with internationally diffused languages, regardless of their development level. Among the top 30 origin countries worldwide in 2010 we find many European states, starting with the UK (with over 1.5million hs emigrants) and Germany (1.25 million), followed by Poland, Italy, France, Russia, the Netherlands, Ukraine, Romania, Greece, and Serbia. In a few cases, this translates into rather high migration rates (17% in Poland, over 20% in Romania, and 11% in the UK) or at least in above world-average ones (9% for Germany, 8% in Italy and 6% in France).

At the same time, European countries have the lowest share of hs immigrants over total immigrants. This is due to a combination of immigration policies, which usually do not select by skill, and geographical or historical factors⁸. The net result of hs emigration and

⁷DIOC stands for Database on Immigrants in OECD Countries, as the original database reported information only on immigrants to OECD countries. However, the most recent release (DIOC-E) and extensions in Artuç et al. (2015) also include information on immigration to non-OECD countries. For DIOC methodology, see Widmaier & Dumont (2011) and Arslan et al. (2015); data are downloadable from: <http://www.oecd.org/els/mig/dioc.htm>. For related data, see Frederic Docquier's website (<http://perso.uclouvain.be/frederic.docquier/oxlight.htm>).

⁸Selective immigration policies are those that target specifically hs migrants. They mainly consist in allowing tertiary educated immigrants to enter the country even before having found an occupation, and/or in reserving quotas for specific professional figures. The most notable cases are those of Australia, New Zealand, and Canada. For a comprehensive discussion, see Chaloff & Lemaitre (2009). For a critique see Belot & Hatton (2012), who show that geographic and historical factors (such as physical proximity or former colonial ties between origin and destination countries, which reduce self-selectivity of migrants) might affect the skill composition of migrants as much as policies. Bertoli et al. (2016) point out that migrants' skill may not be entirely captured by education and other observable characteristics, and that selection based only on the latter can lead to perverse effects on overall migrants' quality.

immigration is nonetheless positive for the largest and most developed, including the UK, France, and Germany. Italy is the main exception, as it suffers of a net loss. As for the least developed European countries, they suffer both of net losses, and high brain drain rates, the main exception being Russia. Notice that non-European large contributors to his migration, such as India and China, do not suffer of any brain drain. This is due to the size of their population and its generally high education level.

Among the most important categories of his migrants, we find doctoral degree holders, especially in scientific and technical fields. DIOC data do not include separate figures for them, but some information can be obtained from the survey on the Careers of Doctorate Holders (CDH), conducted jointly by the OECD and UNESCO in 2007 and covering 25 OECD countries (plus a seven-country pilot project in 2003; see (Auriol, 2007, 2010). Although not explicitly targeted at migration, and even less at innovation, the CDH dataset contains useful complementary information to his migration statistics. First, we learn that “the labour market of doctorate holders is ... more internationalized than that of other tertiary-level graduates” [(Auriol, 2010); p.19]. Immigration rates for doctorate holders are often double those of tertiary educated in general, ranging from 13% in Germany to 42% in Switzerland (Auriol, 2007).

France, Germany and the UK emerge as the most important destination countries along with the US, but most of the international mobility in Europe takes place within the continent, while the US is the top destination for doctorate holders migrating from East Asia and India (who make 57% of foreign doctorate holders, as opposed to only 27% of Europeans).

Another important category of his migrants are academic scientists, which are the object of the GlobSci survey (Franzoni et al., 2012; Scellato et al., 2015). The survey concerns authors of papers published in high quality scientific journals in 2009, who appear to be active in the 16 top countries for number of papers published. Foreign-born authors (defined as those who entered the country of affiliation after the 18th year of age) are more than half of all authors in Switzerland (57%) and around a third in the US (38%). They are in between a third and a fifth in several European countries. The only top countries with limited foreign presence are Spain (7%), Japan (5%), and Italy (3%). GlobSci also confirms that migration within Europe is mainly intra-continental and driven by proximity and language effects; and that the US are the main attractors of Chinese and Indian nationals.

1.3 Migration and innovation in destination countries

In very recent years, various attempts have been made to exploit archival data for retrieving information on the impact of immigrants on their destination country, almost all of them centred on the US. The United States have been an historical destination for foreign-born scientists and engineers, with universities playing a key role in encouraging the inflow of foreign students and postdocs, the former now making around 45% of graduate students enrolled in S&E programmes, and around 60% of postdocs (2006 data, as reported by Black & Stephan (2010)). A debate is ongoing in both the US academic and non-academic press on the extent of foreign researchers' contribution to scientific advancement and innovation, and the related visa policies to undertake. Concerns about the possibility of local S&E students and workers being crowded out have been expressed by several migration scholars, such as Borjas (2009). Evidence in this sense is the dramatic drop of US citizens' enrolment in S&E university programmes, or their marginalisation in some scientific disciplines (Borjas & Doran, 2012). In addition, it has been noticed that more recent cohorts of foreign-born academic researchers in the US tend to concentrate in more peripheral and less productive universities and departments, which do not offer attractive career prospects to native students (Stephan, 2012; Su, 2012). And yet, such evidence could simply prove the existence of a division of labour based on comparative advantages, with US citizens entering professions for which mastering the local language and culture, as well as having a larger social capital, matters more than having acquired specific scientific or technical skills. Several efforts have been made, therefore, to investigate whether absolute advantages may be at play, such as when migrant S&Es are self-selected on the basis of superior skills, thus bringing with themselves knowledge assets and skills that would be otherwise unavailable. Overall, the evidence is in favour of this second hypothesis.

Stephan & Levin (2001) pioneer study focus on the presence of foreign-born and foreign-educated among eminent scientists and inventors active in the US in 1980 and 1990. The authors assemble a sample of about 5,000 highly productive or distinguished S&Es, which include members of the National Academy of Science (NAS), the National Academy of Engineering (NAE), the authors of highly cited scientific paper, a selection of academic entrepreneurs in the life sciences, and a small number (around 180) of inventors of highly cited USPTO patents. The share of foreign-born and that of foreign-educated individuals in each of these categories is then compared to the equivalent shares in the US S&E labour force,

the latter being calculated on the basis of NSCG data (National Survey of College Graduates). Two-tail Chi-square tests prove that in all cases but one the foreign-born are over-represented in the eminent scientist and innovator group. In a few cases, a cohort effect is detected, with foreign-born entered in the US before 1945 being particularly productive (this is not the case, however, for top inventors and academic entrepreneurs). Finally, the foreign-educated are found to contribute disproportionately to these results, which suggests both that the US benefit of positive externalities generated by foreign countries and that immigrant S&Es are self-selected on the basis of skill. Recent work by No & Walsh (2010) confirm this evidence, at least for inventors.

Stephan's and Levin's results on the contribution by foreign-born to entrepreneurship are confirmed for more recent years by a number of surveys conducted by Wadhwa et al. (2007a, 2007b). The authors find that around 25% of all engineering and technology companies established in the U.S. between 1995 and 2005 were founded or co-founded by at least one foreign-born. The percentage increases remarkably in high-tech clusters such as the Silicon Valley (52%) or New York City (44%). These foreign entrepreneurs are mostly found to hold doctoral degrees in S&E, and to be better educated than control groups of natives.

Immigrants' exceptional contribution to patenting has been further confirmed, for a large sample of college graduates, by Hunt (2009, 2013) and Hunt & Gauthier-Loiselle (2010). This depends chiefly on a composition effect, the foreign-born graduates being more likely to belong to S&E disciplines. In addition, it is confirmed that the foreign-born graduates who hold an advantage over natives got their PhD in their country of origin. However, Hunt (2009, 2013) shows that engineering and computer science graduates from the least developed countries face difficulties in getting an engineering job or in reaching managerial positions, being impeded by lack of language skills or social capital. On the contrary, immigrants from richer countries or Anglophone ones (such as India) are more common among the foreign-born actually working as engineers.

Chellaraj et al. (2008) make use of a production function approach to estimate the impact of both foreign-born workers and international students on innovation in the US. The elasticity of patents to the presence of skilled immigrants is found to be positive and significant, and even more so the elasticity with respect to foreign graduate students. This difference can be explained with the composition effect we discussed above: while hs

immigrants comprise many professions, foreign graduate students in the US are concentrated in science and engineering and therefore have a much more direct impact on innovation.⁹

A partial exception to the US-centrism of the literature examined so far is the study by Niebuhr (2010), which focuses on cultural diversity (proxied by the share of foreign born) in R&D employment, as opposed to total employment, as well as in other professions classified as hs. She then investigates the effect of cultural diversity on the patenting rate of 95 German regions over two years (1995 and 1997), finding a positive association. Other studies on the impact of cultural diversity on innovation and growth are those by Ottaviano & Peri (2006), Ozgen et al. (2011) and Bellini et al. (2013), all based on an innovation production function approach, and Nathan (2015).

A growing number of papers exploits new techniques of ethnicity identification based on the analysis of names and surnames, which can be applied to archival data. The most comprehensive enquiry based on this technique has been conducted for the US by William Kerr, in a series of papers based on the NBER USPTO Patent Data File (Jaffe et al., 2001).

Descriptive analysis by Kerr (2007) reveals several stylized facts, most of which are coherent with those concerning hs and scientific migration:

- (i) The ethnic inventors' share of all US-resident inventors grows remarkably over time, from around 17% in the late 1970s to 29% in the early 2000s, that is in the same order of magnitude of CDH estimates of the foreign-born share of doctoral holders, but in a much larger one than that for hs migration from DIOC.
- (ii) The fastest growing ethnic inventor groups are the Chinese and Indian ones, while the overall growth appears to be stronger in science-based and high-tech patent classes.
- (iii) When distinguishing patents according to the institutional nature of the applicant (academic vs. business) it appears that the growth of ethnic inventorship occurred early on in universities, with firms catching up later (in coincidence with the rise of the phenomenon of ethnic entrepreneurship described above).

⁹In a related paper, Stuen et al. (2012) examine the impact of foreign-born (by origin country) vs. native students on the scientific publications (number and citations received) by 2300 US university departments. Foreign-born and local students are found to impact similarly on their departments' publication activity and quality, which goes in the direction of suggesting their substitutability.

(iv) Ethnic inventors appear to cluster in metropolitan areas (with a correlation between city size and percentage of ethnic patents), thus contributing to the growing spatial concentration of inventive activity observed in the US over the past 20 years (this evidence is reprised in detail by Kerr (2009)).

Based on the same technique and on patent applications filed at the European Patent Office (EPO), Breschi et al. (2014) identify inventors of foreign origin also in several R&D intensive European countries (plus Japan and South Korea). The authors then run a series of regression aimed at testing the hypothesis of foreign inventors' superior productivity, as measured both in terms of number of patents and citations received by patents. Their findings confirm the overall positive impacts of inventors of foreign origin on host countries innovation and the skill-based self-selection of these inventors. However, while the results for US and Europe as a whole appear similar, the results appear more nuanced when examining individual European countries. In small European countries, such as Sweden and Switzerland, where the presence of foreign inventors is massive, no strong self-selection effect is detectable. The same applies to Italy, for opposite reasons (the country is generally unattractive for R&D workers, witness the very low number of foreign inventors). As for France, Germany, and the Netherlands, not all entry cohorts appear to exhibit superior productivity. This suggests that other forces besides self-selection are at work that push foreign inventors to immigrate, such as political shocks (for example, the end of restrictions to emigration from former Soviet-block countries in the early 1990s) or family reasons). The only European country whose patterns are very similar to those of the US is the UK.

1.4 Migration and innovation in origin countries: “*Knowledge remittances*” from hs migrants

A longstanding tradition of emigration studies has consisted in evaluating the type and extent of positive returns from emigration for origin countries. Early studies placed special emphasis on emigrants' financial remittances and the role they might play in capital formation in less favoured countries and regions. However, with the increasing importance of hs migration, the emphasis has shifted migrants' contribution to knowledge formation and innovation. These “knowledge remittances” may come in three, not mutually exclusive forms, namely:

- (i) *“Ethnic-bound” knowledge spillovers.* Emigrant scientists and engineers may retain social contacts with former fellow students or educational institutions in their home countries, and transmit them the scientific and technical skills they have acquired abroad (either on a friendly or contractual basis, through visiting professor programmes, research collaborations, or firm consultancy).
- (ii) *Returnees’ direct contribution.* Emigrant scientists and engineers who have worked as academic or industrial researchers, may decide to move back to their origin countries and continue their activities over there. In the case of entrepreneurs, they may keep base in the destination countries, but set up new or subsidiary companies in their home country (Kenney et al., 2013; Meyer, 2001; Wadhwa et al., 2009a, 2009b).
- (iii) *Diaspora networks.* Emigrant scientists and engineers working abroad may decide to come together as an associative platform, this in order to establish collaborative links with their respective home countries or regions. The main goal of such networks is to channel knowledge back home. Most of them are formed by his migrants from Emerging and Developing countries. Some of these networks are even supported or initiated with the collaboration of sending countries governments, while others perform under an informal setting and yet fulfil the original intent which is knowledge exchange (Pyka, 1997).

While case studies on these phenomena abound, large scale quantitative evidence is scant, and largely based on patent data and (mostly) name analysis techniques. The earliest patent-based contribution comes from Kerr (2008), who measures knowledge flows with citations running from patents filed at USPTO from outside the US (in years 1985-97) to patents filed up to ten years before by US residents¹⁰. Citations are grouped according to four criteria (inventor’s ethnicity and technological class of the citing patent, plus inventor’s ethnicity and technological class of the cited patent). Co-ethnic citation groups (in which the country of

¹⁰The use of patent citations to measure knowledge flows is both widespread and controversial. It originates with Jaffe et al. (1993)’s application to the theme of spatial concentration of knowledge spillovers, where it is proved that citations are more likely to occur between patents by co-localized inventors, after controlling for the spatial concentration of patents, by technological classes. This exercise has been criticized for methodological reasons by Thompson & Fox-Kean (2005). Breschi & Lissoni (2005, 2009) and Agrawal et al. (2006b) prove that other types of distance between inventors, in particular social distance, matter as much or more than spatial distance. Technical issues are reviewed by Breschi & Lissoni (2005). For a general critique of the use of patent citations in innovation studies see Roach & Cohen (2013).

origin of the citation and the cited inventor coincide) are found to be on average 50% more numerous than mixed ones.

Kerr (2008) further uses patent data as regressors in a first-difference panel data econometric exercise concerning origin countries of immigration into the US. Here the dependent variables are alternative measures of economic growth. He finds that a one percentage point increase of ethnic patents in the US is associated to a 10% to 30% increase of the country of origin's output measures. The result weakens, but resists, when excluding China from the origin country set, or Computer and Drugs from the technologies considered. This suggests that ethnic-mediated spillovers, while having a stronger impact in high technologies and in one particular economy, are not irrelevant for a more general set of countries and technological fields.

Kerr (2008) pioneer finding can be interpreted by recalling the vast economic and sociological literature on the tacit nature of technical knowledge and the roles of social ties (Breschi & Lissoni, 2009; Jaffe et al., 1993; Miguelez, 2016). A straightforward application of this perspective consider co-ethnicity as a social bond between inventors, which favours the transfer of tacit knowledge assets not comprised in the patent description, but necessary to the understanding and development of the invention.

Agrawal et al. (2008) explore this possibility by assembling a database of ethnic Indian inventors of USPTO patents, all residents in the US in between 1981 and 2000, and find evidence of inventors' co-ethnicity to increase the probability to observe a patent citation, especially for inventors who are not co-located in space. However, when the same authors extend the analysis to citations flowing from India to the US, the evidence for a role of ethnic ties considerably weakens, with the only, partial exception of Electronic technologies. Alnuaimi et al. (2012) and Almeida et al. (2010) also do not find strong evidence in this direction.

Breschi et al. (2015) extend this evidence to migrant inventors active in the US from other Asian countries of origin besides Indian as well as several European ones. In particular, they find that "diaspora" effects do not necessarily translate into "brain gain" effects for the migrants' countries of origin. Namely, while migrant inventors from the same country of origin may have some advantage in accessing the knowledge produce by their fellow migrants, this does not necessarily translate into an easier access to the same knowledge by inventors active in the home country.

In a similar fashion, Foley & Kerr (2013) exploit Kerr (2008) database to investigate the specific role of ethnic inventors in relation to multinational companies' (MNCs) activities in origin countries. In particular, they find that US multinationals with a high share or quantity of migrant inventor's patents invest and innovate more in their ethnic inventors' origin countries, while at the same time relying less on joint ventures with local companies for doing so. This suggests that migrant inventors, act as a substitute of local intermediaries, thus diminishing their companies' costs of engaging into foreign direct investments.

Miguelez (2016) studies how migrant inventors in industrialized countries help increasing the internationalization of inventive activity in their countries of origin. Migrant inventors are identified on the basis of information on their nationality, as provided by patent applications filed according to the Patent Cooperation Treaty (PCT) procedure from the period 1995 to 2011¹¹. Using a gravity model, the author finds evidence of a strong and positive relationship between international collaboration in patenting activities between pairs of developing-developed nations and the stock of migrant inventors from that particular developing nation living in the host developed nation. More precisely, a 10% increase in a the inventor diaspora in a given high income economy leads to an average 2,1% increase in international patent co-inventorship between that economy and the home economy.

Studying knowledge remittances through patent data has several limitations. Patents tend to be more representative of knowledge fields whose absorption requires extensive R&D efforts. In several countries of origin of migrants, just a few or no firms can afford investing heavily in R&D and produce non-trivial numbers of patents. In the same countries, however, public funding of scientific research may sustain the publishing activity of several academic researchers, which suggests that publication data could better capture knowledge absorption through migration (Velema, 2012). Publication-based studies, however, are still very few. An interesting piece of evidence of foreign-born scientists' high propensity to collaborate with colleagues from their home countries and fellow expatriates is provided by Scellato et al. (2015) from a study based on the GlobSci survey. This study suggests that, at least within academic science, some "ethnic" networks are at work, with the potential of delivering knowledge spillovers to origin countries. This means these networks are connected by linkages which expand far beyond host countries boundaries to home countries. Based on a survey of 497 foreign researchers in Italy and Portugal, Baruffaldi & Landoni (2012) find that the probability of productivity and of returning home is high for researchers that maintain

¹¹ A detailed description of the data can be found in Miguelez & Fink (2013).

linkages with home. And even after returning home, researchers still carry on their scientific activities and keep their ties in their social and ethnic networks (Jonkers & Cruz-Castro, 2013; Jonkers & Tijssen, 2008).

A number of qualitative contributions provide more in-depth descriptions of the role played by diaspora network in home countries' innovation (Barré et al., 2003; Brown, 2002; Kuznetsov, 2006). Meyer & Wattiaux (2006) identify 158 networks of hs immigrants from developing countries worldwide. One striking outcome from these qualitative studies is that the successful impact of diaspora networks in their homelands has a lot to do with their direct and indirect participation to their home countries' innovation activities through technology and skills/knowledge exchange platforms, such as innovation fairs, periodic summits, conferences and workshops held in their homelands (Adepoju et al., 2008). The indirect and part of the direct contribution of hs diaspora is not necessarily captured by patent data. This might be an explanation to the contrast between findings by studies on patent citations, and those on publication and collaborations. Furthermore, while returnees and hs diaspora members in origin countries do not always act as a direct source of knowledge transfer, they may still support foreign investments in their home countries through indirect activities such as reference, advice and brokerage. This is particularly true for the case of intra-company mobility within Multi-national Corporations (MNCs) which still remains a grey area in the literature on the mechanisms linking hs migration to innovation.

1.5 International mobility of hs workers within companies and innovation

Most of the literature we have reviewed so far have focuses on specific categories of hs migrants, namely researchers, academics and inventors. However, there are other groups of migrants that might play a role in cross-border knowledge diffusion. These are the professionals contributing to international, intra-company mobility, also known as business migrants or expatriates. Indeed, transfer of knowledge or skill either within the company or beyond it is usually the main motivation of intra-company mobility. And in this case, the nature rather than the level of skills seems to matter the most, in so far that the group of intra-company migrant workers encompasses a myriad of professionals and education levels – executives, managers, engineers, technicians... – as compared to the more homogeneous

groups of researchers, academics or inventors. Migration scholars have not yet devoted much attention on the mechanisms of international mobility of workers within MNCs (Bozkurt, 2006). And there is not much said on how the dynamics resulting from such mechanisms translates into innovation per se¹². This is partly due to the fact that the topic of intra-company international mobility of workers fits into a wide range of disciplines – the sociology of migrations, human geography, the theory of organizations, international economics and human resources management (Peixoto, 2006). It therefore becomes more challenging to find the right approach for tackling it. Furthermore, the intra-company phenomena seems to be more complex and dynamic than the spontaneous or strategic decision of a migrant who chooses one destination country over another with respect to various factors, and as such whose movement is easier to capture. Here instead, the direction of knowledge flow might be much more difficult to identify since MNCs workers could be sent abroad to a company's branch, either from their home country or from another host country with the mission of fulfilling the MNC's assignment.

In general, the literature on the topic of intra-company migrants and their contribution to knowledge diffusion often goes outside the realm of economics of migration or economics of innovation but rather fits into management field of studies (Bjorkman et al., 2004; Bonache & Brewster, 2001). Besides, most existing studies on this question have addressed it under a firm perspective at the expenses of a workforce approach centred on migrants themselves. Indeed, the latter appear to be the main actors of intra-company mobility since they carry with them skills/knowledge that require face to face interaction for transmission – tacit knowledge – and as such these skills/knowledge can be subject to failure or fed with distance (Rodriguez, 2002). In these studies, the key topics have rather been centred on the management of knowledge transfer from MNCs headquarters to their subsidiaries or within 'transnational communities', in other words the social network – spurred by inter-personal relationships – developed within the company and outside national boundaries (Beaverstock, 2004; Lowendahl et al., 2001; Morgan, 2001). There are however few studies with interesting findings which are somehow relevant to the question of migration and innovation within MNCs. For instance, Ozgen et al. (2013) find cultural and past experience diversity of top

¹² There is an increasing "virtual mobility" as a consequence of the progress made in the New Information, and Communication Technologies (NICT), altogether with higher frequencies of business trips which contribute to the complexity of mobility patterns (Collings & Scullion, 2009; Peixoto, 1999; Salt, 2008; Salt & Wood, 2012) and so the difficulty to veritably capture information exchanged– amongst MNC hs mobile workers on one side and amongst them and locals in the subsidiaries on the other side – and consequently its relationship to innovation. Also, the relatively low duration characterizing such intra-company international movements has contributed of making some migration scholars less reluctant to qualify them as migration flows.

management team (TMT) to be positively associated with firm innovation. This implies experience accumulated by migrants and their interaction with locals and as individual from different backgrounds could be a great asset to the firm. Similarly, a positive impact of TMT cultural diversity has been found on firm performance and the internationalization of firm's activities (Greve et al., 2009; Heijltjes et al., 2003). One could explain this limited amount of studies with the small size in absolute terms of intra-company international flows. Although with the globalisation of the world economy, there has been an increasing frequency of intra-company international mobility. Another reason for this lack of interest could be the absence of systematic large-scale data on this category of migrants. The very few existing studies have relied on surveys for data collection – this comes with all the limitations attached to data collected with this method, such as data infrequency and relatively small size of surveyed samples. There are however striking observations emerging from some of these studies.

The first of these observations is that MNC workers migrating to developed countries from developing ones within the same company often respond to scarcity of local specialists in a given field in these developed countries. This scarcity of local specialists could be explained by the comparative advantage local workers might be having in following other specialities or entering other professions that demand a more thorough socio-cultural local background than having acquired specific technical skills, as already mentioned in section 3 for the case of US citizens. These migrants belong to the category of migrants which has been referred to as 'International Staff' by Perkins (1997) and are often the target of many selective immigration policies in developed countries. Indeed, foreseeing the positive impact migrants with a good level of skills in specific fields could have on their economy, many developed nations have adopted specific migration policies in order to ease their inflow through this channel. The 'Green Card' introduced by the German federal government in 2000 for instance has helped the country to cope with increasing demand of specialists particularly in the IT sector (Burkert et al., 2008). But the biggest winners of this program were probably German MNCs as they were encouraged to fill skills shortage by transferring IT specialists from their branches in developing countries to their headquarters in Germany. (Meijering & van Hoven, 2003) present an interesting development and evidence on the experience of 22 Indian IT professionals who migrated to Germany through their companies' internal transferring. Similar analysis was conducted by Burgers & Touburg (2013) for the case of a group of Indian IT professionals based in the Netherlands, and that were recruited internally from India within the French Capgemini. In both case studies the authors found that their recruitment was mainly motivated by a shortage of highly qualified local software specialists.

Another observation that can be drawn from the literature is that unlike the intra-company mobility of workers from developing to developed countries which responds to a given demand in a specific field, the intra-company mobility of workers from developed to developing countries is motivated by a more accurate vision of skills or expertise transfer to the firm's subsidiaries at the destination country. In this case, expressions such as 'expatriates' (Tan & Mahoney, 2006) or 'Parent/Headquartered Staff' (Perkins, 1997) will be commonly used in the literature for migrants, in reference to the direction of flows which usually goes from the parent companies or headquarters to their branches. MNCs use different approaches to the international management of knowledge transfer on which depend partially or entirely successful assimilation of this knowledge by locals within the company framework. One of these approaches which has been shown to be successful is related to cases where skills and expertise of managers from MNCs headquarters are only used at the very initial phases of a project implementation in developing countries subsidiaries. With the help of on-the-job trainings and face-to-face meetings, those skills and expertise are then passed on to local employees (Martins, 2011). This appropriation of skills and expertise by locals is what constitutes one of the main difference between intra-company mobility of hs workers from developed to developing countries and intra-company mobility from the opposite direction. Indeed as noted above, local workers from developed countries might deliberately choose not to acquire certain skills for career development perspectives while for local workers of developing countries this lack of skills might be a matter of lack of resources rather than a matter of choice. There are however other cases where MNCs heavily transfer parent country nationals to their developing country branches in order to gain more control over these subsidiaries. This is likely to result to low communication and exchange with local employees who are often kept at very low skilled level jobs. In this vein, this approach could lead to the vicious double effect of transferring some new knowledge to specific locations and at the same time contributing to digging the skills gap between local workers and those at the headquarters. Case studies have shown such practices to be common with Japanese MNCs (Barlett & Yoshihara, 1988; Harzing, 2001; Kopp, 1994).

All in all, although international mobility of hs workers within MNCs often translate into some new knowledge or skills brought in to the functioning of the host headquarters/subsidiaries, there still remains some grey areas supported by mixed evidences on its impact on host developing countries innovative capacity and development in general. Indeed, there have been controversial debates on the ability of expatriates to veritably infer some knowledge diffusion to local firms in developing countries. As pointed out by

McAusland & Kuhn (2011), new ideas or knowledge are mainly relevant in the country of production of these skills/knowledge. This implies skills/knowledge carried by expatriates may not match with the needs and skills of local firms and so will strictly remain within the MNC compound. Admittedly, the first motivation of MNCs when engaging into international transfer of hs workers is quite far away from the local innovation or development goal in host developing nations. They are rather led by an intra-company knowledge dissemination strategy – that can also be perceived as a technical and business security strategy –, designed exclusively to fit into the MNC's process and 'know-how' policies (Galbraith & Edstrom, 1976; Hocking et al., 2004; Thomas, 2008). However by following the above reasoning, one risks narrowing the approach to knowledge diffusion by omitting the possibility of assimilation by local firms of that new knowledge initially designed for MNCs and its adaptation to their needs, provided some sufficient absorptive capacities in these local firms (Ikiara, 2003; Mansfield & Romeo, 1980; Romer, 1996). This is particularly the case when the MNCs R&D is decentralised to the subsidiary level, then there is a great chance to infer some knowledge spillovers to local firms (Sanna-Randaccio & Veugelers, 2007).

1.6 Discussion and conclusions

Once the preserve of research in development economics, the study of highly-skilled (hs) migration has increasingly attracted the interest of innovation scholars. There has been an increasing availability of quality data developed from censuses and labour force surveys from which hs migrants are identified on the basis of their education level. Statistics from these data depict European nations among the largest hs migrants sending countries while they lag far behind the US as hs receiving nations. In addition, innovation scholars have produced more targeted data such as those on academics, doctoral students and inventors, which have allowed to jointly address migration and innovation questions under an empirical setting. Concerning their impact on destination countries, hs migrants have been found to exert a positive effect, as measured by productivity measures, patenting or scientific publications. Although this literature has mainly focussed on the US, there have however been some cross-country studies on European countries which have found similar results.

As for the impact on source countries, the US-centrism of the largest part of the literature has biased the research towards the countries whose hs migration to the US has grown the most in

recent years. By and large, these are China, India, and other East Asian countries, all of which are developing countries. We much less on migration to the US from other developed countries, such as the European ones. This is especially true for inventor-based data. Overall, empirical evidences on inventors indicate that a diaspora network effect exist within receiving countries, with a high propensity of inventors with same ethnicity to collaborate among each others. But evidence on positive spillovers for their origin countries are mixed. Indeed, some studies find positive results such as hs migrants having a high propensity to collaborate with their home country counterparts and/or determining an increase of economic growth rates, inventive activities and MNCs' direct investments in their origin countries. On the contrary other studies, while finding evidence of close ties among migrant inventors from the same origin, do not find that they translate automatically in technology transfers to their home countries. The only study that examines developed origin countries find some evidence of knowledge flows reaching the migrants' origin countries through the MNCs' channel, but not through the social network one.

The diversity of data sources may provide an explanation for this mixed evidence. Yet, patent data remain the most commonly used type of data in this literature, which could be perceived as one limitation to the extent that patent data do not necessarily capture the indirect and part of the direct contribution of hs diaspora. Qualitative studies show the successful impact of hs diaspora networks in their homelands to be mainly attributed to their direct as well as indirect participation to their home countries' innovation activities. Finally, more attention needs to be paid to intra-company migration, more particularly on the international transfer of workers with specific skills within MNCs. Most papers on this topic mainly address questions related to firms' organizational theories and thus best fitting into the area of management studies. Although the first purpose of such intra-company international mobility is to transfer skills/knowledge to the headquarters/subsidiaries, with some externalities to the host economies, this topic remains a grey area in the literature of migration and innovation.

Chapter 2

Immigration externalities, knowledge flows and brain gain

2.1 Introduction

Highly-skilled workers are an important asset for a country's growth as they impact directly on knowledge production and diffusion (Nelson & Phelps, 1966; Vandebussche et al., 2006). The network of knowledge exchanges developing around these highly-skilled workers generates externalities to the actors being part of that network (Breschi & Lissoni, 2009; Lodigiani, 2009). In the literature, a special attention has been given to the international dimension of such networks, with a focus on highly-skilled migrants as the main actors and their contribution to innovation or knowledge transfer to their host countries (Breschi et al., 2014; Kerr, 2008; Stephan & Levin, 2001), on one hand, and to their homelands, on the other hand (Brinkerhoff, 2006; Kapur, 2001; Kuznetsov, 2006; Saxenian, 1999). Highly-skilled migration has been for long considered as a "lost" of human capital or a brain drain to the home countries, while there has been studies documenting this brain drain to be beneficial to the host countries which get this brain intake, or brain gain. In this context, a number of case studies have been produced, which looks into whether substantial knowledge "remittances" may compensate for the loss to the sending countries. However, systematic empirical evidence is still scarce, with few recent works looking at the role highly-skilled migrants play on the diffusion of knowledge from the US to their homelands (Agrawal et al., 2011; Kerr, 2008). In particular, no study has yet tested for the exact nature of the link between highly-skilled migrants and knowledge flows to their home economies at a more global scale. As for the impact on highly-skilled migration to the host countries, while there have been numerous studies conducted around other specific aspects of that impact – such as impact on the labor market, on productivity or economic growth –, the knowledge flow or innovation aspect has not received as much tribune in the literature; with the exception of few studies in the US (and lately also few European countries).

This paper intends to fill this gap of knowledge. In particular, it explores knowledge feedbacks to home countries and knowledge inflows to host countries, both generated by migrant inventors, a representative category of highly-skilled migrants - most of them scientists and engineers. We test our hypotheses of a positive relationship between knowledge flow and highly-skilled migration in a gravity model framework. For doing so, we look for how highly-skilled migrants – our focal regressor –affects knowledge flow to home and host countries, as measured by patent citations – our dependent variable. The present empirical analysis is made possible by the use of a novel dataset of inventors with migratory background as a proxy for a highly-skilled migrants (Miguelez & Fink, 2013).

Results from our regressions confirm our initial assumption on the positive impact of highly-skilled migrants on knowledge flow, not only to their host countries but also to their homelands. More precisely, we find that doubling the number of inventors of a given nationality at a destination country leads to 8.3% knowledge flow to their home economy from that same host land. A similar increase in the number of migrant inventors produces a 6% increase in the knowledge inflows to the host economy. We also find that these results are not driven neither by the US as a traditional country of highly-skilled immigration, nor by India and China as the biggest highly-skilled sending countries, since we obtain very similar results even after having dropped these three countries.

The rest of the paper is organized as follows: the next section presents selected literature on highly-skilled migration and innovation. Section 3 focuses on the research method, including the description of our data and variables, a presentation of the model. A discussion of our results is done in section 4. We conclude in section 5.

2.2 Background literature

Innovation can be viewed as an outcome of cumulative efforts in R&D in the domestic country or region. Early empirical evidence on international technological diffusion have shown a country's R&D capital stock – a measurement for the stock of knowledge – could positively impact on its own total factor productivity (Coe & Moghadam, 1993; Griliches, 1988), but also on the one of its trade partners (Coe & Helpman, 1995). However, Keller (1997) casts some doubts on Coe & Helpman's (1995) findings showing that similar results were obtained even when using random numbers for trade patterns. In a further analysis, other scholars (Eaton & Kortum, 1996; Keller, 2001) suggest technological diffusion could as well

be a consequence of other regional effects not linked to trade or could come from the simultaneous effect of several other channels like geographical distance, Foreign Direct Investment (FDI) and language similarity. Yet, Coe & Helpman (1995) study provides an early empirical evidence on the determinants of the international flows of knowledge or technological diffusion. In general in the innovation literature, variables such as trade, R&D and FDI have been used to proxy the extent of knowledge exchange between countries. However, finding a good measurement for knowledge flows could be cumbersome to the extent these flows are not tangible and as such it becomes a hard task to find a comprehensive instrument for capturing them. The use of patent citations emerged as a way of overcoming this limitation. This technique was pioneered by Jaffe et al. (1993). Indeed, using patent citations from the United States Patent and Trademark Office (USPTO), Jaffe et al. (1993) investigate patterns of technology diffusion in the US. From geo-localising patents on the basis of their inventors' addresses, they find a higher likelihood for US patents to be cited by other patents coming from the same state than by patents from other states, thus pointing out to the existence of localized knowledge flows. Since then, this technique has been widely applied to various other studies – as no better measurement of knowledge flows or technology diffusion has been found up to date – which have attempted to investigate channels of knowledge flows and what drives such flows at a cross-border level.

One of these studies is conducted by MacGarvie (2005) in which she applies a cross-country gravity model¹³ using citation data from the USPTO in order to test for the determinants of knowledge flows. The author takes the number of patent citations between country pairs as the dependent variable and a proxy to knowledge flows, among ten large countries (for the period 1980–1995). The author then includes some dyadic control variables such as the technological and cultural proximity – a dummy for common language, geographical distance between pair of countries, citing countries' imports from cited countries, as well as country's specific variables like citing countries' employment of R&D workers, FDI and the yearly number of telephone calls. The author finds FDI, geographical, technological and cultural proximity favors knowledge flows between two countries. Meanwhile, trade is associated with knowledge flows only in inventions within the same fields of technology.

¹³ The gravity model of migration is a model in urban geography derived from Newton's law of gravity, and used to predict the degree of interaction between two places (Rodrigue et al., 2009). In other words, it is a prediction model borrowed from physical gravity by social sciences, and economics in particular, which basically seeks to explain the interaction between two elements with the characteristics of each of these two elements and the distance between them.

The international mobility of human capital is gaining attention in the related literature as a channel of international knowledge diffusion. In general, studies on highly-skilled migration and innovation has been for a long time confined to the area of migration and development (De Haas, 2010; Docquier & Rapoport, 2012).¹⁴ However, a group of scholars have worked on linking it to innovation studies mainly with the help of patent data (Agrawal et al., 2011; Breschi et al., 2017; Kerr, 2008; Miguelez, 2016).

There is indeed a well-established theoretical literature on the brain drain phenomenon (Beine et al., 2001; Berry & Soligo, 1969; Bhagwati, 1975; Grubel & Scott, 1966; Johnson, 1967; Lucas, 1990; Mountford, 1997). At a first glance, the conventional migration literature seems to depict brain drain mainly as the concern for developing nations or for south-north migration. This could be explained by developing nations' relatively low endowment in human capital as compared with other nations' groups. It thus makes them more likely to be vulnerable to any human capital loss, particularly a loss of their highly qualified one. This explains why high income countries are often cited among the winners in the brain drain debate, in so far that their high endowments in human capital seems to make the impact of the massive highly-skilled emigration they experience negligible in relative terms.

However, if we consider the basic definition of brain drain to be the permanent emigration of qualified workers¹⁵ (Straubhaar, 2000), it should be acknowledged that brain drain is not only restricted to developing countries as high income nations such as Germany, France and the United Kingdom have to face a considerably high rate of emigration of their highly-skilled workers. Indeed, from a data on emigration rates by country group in 1990 and 2000, Docquier et al. (2007) report the proportion of highly-skilled workers among migrants to be much higher than the proportion of highly-skilled workers among residents, this for every

¹⁴There has not been yet any general and commonly agreed framework for the definition of a migrant. There are three dimensions that have been used in the literature to define a migrant: nationality, country of birth and duration of stay. In empirical studies, migrant definition seems to be mainly driven by information provided in the used data. Each of these dimensions has its own limitations (for a detailed discussion see Champion (1994)). However the duration of stay dimension seems to be one important element for distinguishing between simple mobility and migration. That is probably one of the reasons justifying the recent emergence of the use of 'diasporas' as a new terminology for qualifying people of same origin, living in one or many places abroad. This new terminology seems to be more flexible in its conceptual ground and in the different dimensions embedded in the definition of a migrant. Diaspora is thus defined as "part of a people, dispersed in one or more countries other than its homeland, that maintains a feeling of transnational community among a people and its homeland" (Chander, 2001).

¹⁵This definition is in contrast with the definition proposed by the United Nations (UN), which depicts brain drain as a one-way movement, that only covers South-North or developing-developed countries migratory flows and only benefit to the latter countries (Gaillard & Gaillard, 1997). The UN definition is somehow misleading and today would sound obsolete in so far that the world competition for innovation has led to an increasing north-north brain drain.

income groups. Furthermore, many high income countries have adopted local policies aiming at keeping their talents at home and attracting foreign ones. Therefore, it appears that all countries share a common concern with respect to brain drain. It is likely that the rate of highly-skilled migration will be maintained over the upcoming years (see Artuç et al. (2015) for evidence from the DIOC data, the Database on Immigrants in OECD Countries). Thus instead of focusing on questions such as which countries are the biggest losers or winners of brain drain, much attention should probably be given to finding strategies in order for sending as well as receiving countries to benefit from highly-skilled migration. This is partly what has been at the core of a new trend of studies that has emerged during the past two decades (pioneered by Stark et al. (1997, 1998) and Stark & Wang (2002)) and which advocates the possibility of some brain gain for both sending and receiving countries from highly-skilled migration.

The benefits from highly-skilled migration to receiving countries are supported by evidence on their over-representation in the host country highly-skilled population and their positive impact on knowledge productivity, thus innovation (No & Walsh, 2010; Stephan & Levin, 2001). Additionally, highly-skilled migrants also strongly contribute to entrepreneurship (Wadhwa et al., 2007) and patenting (Chellaraj et al., 2008; Hunt & Gauthier-Loiselle, 2010) in their host countries (see also Breschi et al. (2014)). One of the main channels of highly-skilled migration to the receiving countries – particularly high income ones – is through higher education. Indeed, universities have been playing a major role in attracting foreign students and postdocs mainly in scientific and technology fields. A study by Black & Stephan (2010) reports foreign students to be accounting for around 45% of graduate students enrolled in science and engineering programs and the share of foreign postdocs to be 60% of all postdocs, this from a 2006 data from the US. An important share of these foreign students and postdocs remain in their host countries after completion of their studies, yet some still maintain ties with their home countries.

While all of the above-cited positive impacts might appear as an evident highly-skilled migration outcome to the receiving countries, it seems not to be the case to the sending countries. However, a different approach to highly-skilled migration has led the recent empirical migration literature to uncover several ways through which brain gain could occur to sending countries. One of these mediums is returnee migrants. For instance, Chacko (2007) provides evidence of the important role of Indian professional immigrants who returned

home, in the development of the IT industry in the city of Bangalore. Other studies have reported higher intentions to return among various groups of skilled migrants (Bollard et al., 2011; Gibson & McKenzie, 2012; Kangasniemi et al., 2007). Another approach to brain gain for sending countries in the empirical literature has been to show how co-ethnic ties eased knowledge flow among inventors of same origin back to the migrants' source country (Agrawal et al., 2008, 2011; Almeida et al., 2010; Kerr, 2008). Using patent data from the USPTO, Kerr (2008) shows ethnic ties to increase knowledge diffusion from the US to the migrants' homelands, as well as technology production in their countries. The effect is even stronger for high-tech industries and for the case of China. The author applies an ethnicity identification technique based on inventor's names – which has inspired several other empirical work such as the one by Breschi et al. (2017).

Besides innovation studies, highly-skilled migration is also the subject matter of international economics scholars. Indeed, migrant networks and ethnic ties have been positively linked to a set of variables in the literature such as Foreign Direct Investment (FDI) (Javorcik et al., 2011; Kugler & Rapoport, 2007), R&D off shoring (Miguelez, 2016) and international trade (Gould, 1994). In these studies, highly-skilled diasporas are depicted as a direct – from diaspora members deliberately interacting with their home economies – and an indirect – from diaspora members serving as an intermediate for easing transactions between host and home economies – vector of economic and knowledge development to their home countries (Kapur & McHale, 2005).

Unfortunately, empirical work on migration and innovation is limited to a group of countries, mainly the high income receiving ones – in particular the US as an historical destination of highly-skilled – and few sending countries like India and China. The focus on these two countries is somehow related to the fact that their highly-skilled migrants make an important part of the total highly-skilled migrants in the US. And this is at the expenses of studies on highly-skilled from other countries like low income countries which are highly-skilled migrant net exporters.

All in all, to the best of our knowledge, very few studies have empirically investigated the actual nature of the link between highly-skilled migration and knowledge flows. While most studies seem to agree on the brain gain for host countries in terms of knowledge flows, this literature has not reached any consensus so far when it comes to the impact on home countries. Some studies have found some positive impacts in terms of knowledge flows as

well as development, while the results from other studies remain inconclusive. This might be due to the specificities attached to samples used altogether with the model specifications, which call for some cautions when generalizing findings from these studies. In the present analysis, we will borrow from the model specifications applied by Miguelez (2016) who investigates the role of highly-skilled in the internationalization of knowledge and the one by MacGarvie (2005) who looks into the determinants of knowledge diffusion. Both papers apply a gravity model using patent data.

2.3 Methodology and data

2.3.1 Empirical approach

For the present analysis we use a standard gravity model.¹⁶ The gravity model in its initial form was used to explain migration patterns (Ravenstein, 1889). However the use of this model in migration related studies has declined throughout decades to slowly be replaced by its application in trade studies as pioneered by Tinbergen (1962). From then on it has been extensively used and extended in several other international trade studies.

The knowledge outflows model takes the following form:

$$KF_{ijt} = e^{\beta_0} \cdot mig_{ijt}^{\beta_1} \cdot Z_{ijt}^{\gamma_n} \cdot e^{\tau_i} \cdot e^{\tau_j} \cdot e^{\delta_t} \cdot \varepsilon_{ijt} \quad (2.1)$$

And for knowledge inflows:

$$KF_{jit} = e^{\beta_0} \cdot mig_{ijt}^{\beta_1} \cdot Z_{ijt}^{\gamma_n} \cdot e^{\tau_i} \cdot e^{\tau_j} \cdot e^{\delta_t} \cdot \varepsilon_{ijt} \quad (2.2)$$

Where KF_{ijt} and KF_{jit} are respectively the number of patent citations from migrant inventors home country i citing host country j in the year t and the number of patent citations from host country j citing migrant inventors home country i in the year t , β_1 is our parameter of interest for the bilateral migration variable, mig_{ijt} is the number of active inventors of i nationality residing in country j during the year t , Z_{ijt} is the set of dyadic and country specific control

¹⁶ The possibility to turn international migration data into a bilateral data has opened up to a wider perspective in terms of the set of research questions that can be addressed by scholars. But at the same time, it has constrained them to the application of gravity models – which used to be confined to the domain of trade analysis – in their empirical research, bringing in all the specification and methodological challenges associated with such a model. A recent paper from Beine et al. (2014) discusses more thoroughly about all of the challenges faced when applying gravity models to international migration studies.

variables in year t , τ_i , τ_j and δ_t are country i , country j and time FE respectively. ε_{ijt} stands as the error term.

When applying the gravity model either to trade or migration we might face the issue of the presence of strong skewness in the data distribution with relatively a few high values to the bottom end. The common econometric usage in the related literature when dealing with such issue has been to transform the gravity equation – equation (2.1) and (2.2) in our case – into its logarithmic form – with a normal disturbance term, then estimate it with an Ordinary Least Squares (OLS). However, the latter practice may induce some heteroskedasticity in the error terms, thus some inconsistency of estimation as pointed out by Santos-Silva & Tenreyro (2006). Consequently, the authors recommend estimating the multiplicative form of the model using Poisson pseudo-maximum likelihood (PPML).

Given all what precedes, we choose to apply the PPML regression to the following conditional expectation of equations (2.1) and (2.2), equations (2.3) and (2.4) respectively:

$$E(KF_{ijt} | X_{ijt}) = \exp[\beta_0 + \beta_1 \ln mig_{ijt} + \gamma_n \ln Z_{ijt} + \tau_i + \tau_j + \delta_t + \varepsilon_{ijt}] \quad (2.3)$$

and

$$E(KF_{jit} | X_{jit}) = \exp[\beta_0 + \beta_1 \ln mig_{ijt} + \gamma_n \ln Z_{ijt} + \tau_i + \tau_j + \delta_t + \varepsilon_{ijt}] \quad (2.4)$$

2.3.2 Data

2.3.2.1 Patent citations as a measure of knowledge flows

Our dependent variable is built using cross-country citations to Patent Cooperation Treaty (PCT) patents—the patent database from the World International Patent Office (WIPO). More precisely, we retrieve backward citations to PCT patents – as cited patents – from the OECD Citations database from July 2014 (Webb et al., 2005) and geo-reference both cited and citing patents across all countries. From the initial data, only citing and cited patents with information on inventors and their addresses are selected, for the period 1990-2010.¹⁷ Then we drop national-level citations and only keep international citations. The baseline table observations are at the inventor level – see a sample in Table 2.1. At a first stage we compute the weight of each inventor per citing and cited patents with their share in the total of inventors per patent. Then we make the product of each share of citing and cited inventors –

¹⁷Due to the relatively low quality and consistency of data prior to 1990 and after 2010, we focus on the period 1990-2010

see the variable in the last column of Table 2.1. The knowledge flow variable is the sum of this product by citing-cited pair, for each year as shown in Table 2.2. Thus our dependant variable is a dyadic variable returning the fractional count of backward patent citations from one country to another per year. As we will see in detail from the statistical description section below, our final data coverage is not evenly distributed across country pairs¹⁸.

Table 2.1: Sample of citations at the inventor level

Citing appli. id	Cited appli. id	Year	Citing inv. country	Cited inv. country	Total citing inv.	Total cited inv.	Share Product ⁽¹⁾
6697	16009750	2008	AT	DE	1	1	1
6697	17392033	2008	AT	DE	1	1	1
6697	54487160	2008	AT	US	1	2	0.5
6697	54487160	2008	AT	US	1	2	0.5
6711	47362375	2008	AT	US	4	1	0.25
6711	47362375	2008	AT	US	4	1	0.25
6711	47362375	2008	AT	US	4	1	0.25
6711	47362375	2008	AT	US	4	1	0.25
6727	15631593	2008	CH	DE	2	1	0.5
6727	15631593	2008	DE	DE	2	1	0.5

Notes: (1) Share product = $(1/\text{Total citing inventors}) * (1/\text{Total cited inventors})$

Table 2.2: Knowledge flows variable

Year	Citing country	Cited country	Knowledge flows ⁽¹⁾
2007	AT	DE	813.009
2008	AT	DE	806.478
2007	AT	US	1,301.547
2008	AT	US	1,367.372
2007	CH	DE	1,263.891
2008	CH	DE	1,379.026

Notes: (1) Knowledge Flow = \sum Share product by citing-cited pair for each year

¹⁸ See Miguelez & Fink (2013) for a detailed description of the migrant inventors' dataset.

We distinguish between two types of knowledge flows with respect to the direction of citations:

- Knowledge outflows which is computed from backward citations received by host countries from home countries.
- Knowledge inflows which is built from backward citations received by home countries from host countries.

All of these two types of knowledge flows are computed in the similar way as explained above.

2.3.2.2 The international mobility of inventors

One striking observation that can be outlined from the literature as described in section 2 above is the lack of a common measurement or definition for highly-skilled, although it should be pointed out that skill level definition is often driven by the data in hands. First existing data on highly-skilled migrants referred to occupation for defining and classifying skills (Koser & Salt, 1997), while some countries on the other hand have used the level of paid wages to measure skills. However in most migration studies nowadays, the level of education has been widely used as a measurement for capturing skill level. In these cases, a highly-skilled is defined as anyone holding a tertiary education qualification or that have accumulated a certain amount of years of schooling. Each of the three criteria of skill identification has its own specificities.¹⁹ Some scholars have found a way to reconcile all three of them by working on inventors as highly-skilled, a category which seems to embody all if not most aspects of skill (Breschi et al., 2017; Kerr, 2008; Miguelez, 2016). One advantage of using inventors' data in highly-skilled migration studies is that migrant inventors stand as a more homogenous category of highly-skilled migrants. This is not the case with similar data where skill levels are obtained from education levels.

For the present analysis we make use of an unused dataset on inventors from PCT patent applications (Miguelez & Fink, 2013), from which we are able to identify inventors with a migratory background on the basis of their nationality and place of residence. It should be acknowledged that nationality is not always the best proxy for migrants' origin. Indeed

¹⁹For a detailed discussion see Chaloff & Lemaitre (2009).

relying on it might increase chances of underestimating migrants, which is not the case of other origin proxies such as country of birth, a proxy commonly used in most migration datasets like the DIOC dataset. An advantage of using this dataset is that it is the only international migration dataset where information on inventors' nationalities and residence are provided by inventors themselves.²⁰ It is therefore possible to identify migrant inventors by confronting information on nationality with those on residence. In our view, nationality appears to be a more natural signal of origin than other proxies based on name identification techniques that have recently emerged in the literature (Agrawal et al., 2011; Breschi et al., 2017; Kerr, 2008).

Note also that an additional point that makes the PCT dataset stands one step ahead from other existing datasets from country or regional patent offices is its international dimension. This thus makes it rather more convenient for the present analysis to the extent that other patent datasets will tend to be more biased towards one or another origin/destination country or region. Furthermore, the dataset yearly coverage enables it to be organized into a panel data with information on both patent citations and migrant inventors for a wide range of countries that can be coupled into citing/cited countries or sending/receiving country pairs.

Yearly country data on migrant inventors from the WIPO dataset are the starting point for computing our focal explanatory variable at a sending-receiving country pair level. More precisely, this variable stands as the total number of inventors who are nationals of a country and residing in a given host country for the years 1990 to 2010.

2.3.2.3 Control variables

Drawing from MacGarvie (2005) and Miguelez (2016), we control for geographical distance as well as cultural and historical ties between citing/sending and cited/receiving countries. Two variables are included for the geographical distance: one is a dummy variable for contiguity, taking the value 1 if the two countries share a common border and 0 otherwise. The other one is a variable measuring the distance – in kilometers – between the biggest cities of both countries, weighted by the share of those cities' population in the overall countries'

²⁰The PCT requires patents applicants (or at least one of the patent applicants) to be a national of a PCT Contracting State, thus the requirement of filing applicants' nationalities as well as their residences. And for international applications filed before 16 September 2012, inventors have to be listed as applicants for the purposes of the U.S. designation.

population.²¹ Cultural ties are proxied with a dummy for common language taking the value 1 if both countries share at least one language and 0 otherwise. For controlling for historical ties, we include a dummy taking value zero if there has been a colonial link between both countries and 0 otherwise.²²

Additionally, we include an index of technological similarity between pair of countries in order to control for whether they both share common fields of technological specialization. This index is computed using patents data from the CRIOS – Patstat database – which is derived from the EPO patents database – from 2014 (Coffano & Tarasconi, 2014) and applying the following formula:

$$Tech. similarity_{ij} = \frac{\sum f_{ih}f_{jh}}{(\sum f_{ih}^2 \sum f_{jh}^2)^{1/2}} \quad (2.5)$$

Where f_{ih} stands for the share of patents of technological class h – according to the 30-class reclassification of IPC codes - of country i , and f_{jh} the share of patents of technological class h of country j .²³ Values of the index close to the unity indicate that a given pair of countries are technologically similar, and values close to zero indicate they are technologically far from each other (Jaffe, 1986).

Lastly, we control for each country level of technological capacity with its total number of inventors per year using information from the PCT data. This variable is informative to the extent it is a signal to the size of a country innovation systems, which clearly determines a country’s capacity to collaborate with foreigners, as well as its capacity to attract inventors from abroad or to be a source of migrant inventors.

²¹By weighting the distance by the population, we would like to take into account the fact that it is more likely to have more traffic between cities with bigger population density than cities with a smaller one. This might in turn make the cities closer, no matter how distant in kilometres they are from each other.

²²All of the above control variables come from the ‘*Centre d’Etudes Prospectives et d’Informations Internationales*’ (CEPII). A detailed description of these variables can be found in Mayer & Zignago (2011)

²³This 30-class re-classification of IPC codes was originally proposed by the OST (Observatoire des Sciences et Techniques). For more details see Coffano & Tarasconi (2014).

Table 2.3: Variables definition

Variables	Definition
Knowledge outflows	Fractional count of citations from citing country i to cited country j.
Knowledge inflows	Fractional count of citations from citing country j to cited country i.
Migrant inventors	Total inventors from origin country i residing in destination country j
Technological similarity	Index of technological proximity between countries i and j at a given year, varying from 0 to 1
Contiguity	Dummy variable for taking value 1 if countries i and j share same border
Colony	Dummy variable taking value 1 if country i is a former colony of country j
Common official language	Dummy variable taking value 1 if countries i and j share at least one common national language
Distance	Distance in kilometres between the biggest cities of both countries i and j, weighted by the share of those cities' population in the overall countries' population
# of inventors in country i	Total number of inventors with country i nationality
# of inventors in country j	Total number of inventors with country j nationality

2.3.2.4 Descriptive statistics

For our sample we have selected 141 countries worldwide grouped into pairs and dropped same country pairs as we are interested in international rather than intra-country knowledge flow. In this section, we present some statistical description of our main variables of interest.

Citations and migration corridors

Tables 2.4 and 2.5 present an overview of the top 20 biggest players when it comes to knowledge flows and inventors' migration respectively for the period 2006 to 2010.²⁴

From Table 2.4 we can see that the biggest flows of citations are amongst technology-leading or high income countries, with the US being the largest origin of knowledge mainly to other

²⁴We choose to represent the total flows over the last five years of our sample as those are the years with the highest consistency of records.

high income nations and China.²⁵ Interestingly, there is a good deal of knowledge flowing from Japan to the US, as the two countries appear in the second position of this list. Also, there is a strong reciprocity in the flow of knowledge to the extent that many top knowledge recipient countries are also a source of knowledge to the countries they get knowledge from. Therefore, it would be interesting to see how much migrant inventors play in this reciprocity. Additionally, the knowledge flow distribution is much skewed to the left, with the top 20 country pairs accounting for up to 49% of the share of bilateral citations worldwide. More precisely, the largest flows of knowledge are exchanged and circulate amongst a small group of countries – 13 countries, almost all of them being high income except for China. When dropping high income countries from the list of knowledge recipient countries, China appears as an important knowledge destination, receiving from the US alone around 20% of the share of citations going to low and middle income countries (Table A1 in Appendix A). This table also shows India as a second leading knowledge recipient country. Interestingly, South Africa appears as the only African country in this list which is mainly dominated by Asian countries as knowledge destination and the US as the main knowledge origin country. Again here, there is a skewed distribution of bilateral citations – 67% of knowledge flowing amongst top 20 country pairs – and high income countries remain the top origin of knowledge. Table A2 in Appendix A reports the top 20 bilateral citation flows when focusing on low income countries only as knowledge recipients. The most striking observation from this table is the dominant position of Kenya and North Korea as the biggest knowledge destinations from the low income group. Furthermore, around 66% of the share of citation pairs with low income as citing countries comes from these 20 biggest players. However, there are not many citations going to this group of countries in absolute terms. They all get their knowledge from high income countries.

²⁵The US alone accounts for close to 29% of all origin of knowledge in the entire sample.

Table 2.4: Citations corridors (Total flow of citations for the top 20 country pairs for the period 2006-2010)

Citing country	Cited country	# of citations	Citation share Cum. (%)	Citing high income	Cited high income
Germany	US	91,218.09	7.02	yes	yes
US	Japan	78,205.18	13.04	yes	yes
UK	US	47,177.25	16.67	yes	yes
US	Germany	45,482.01	20.17	yes	yes
France	US	41,102.07	23.33	yes	yes
Germany	Japan	39,408.45	26.36	yes	yes
US	UK	29,331.3	28.62	yes	yes
Canada	US	27,510.2	30.74	yes	yes
Japan	US	24,573.1	32.63	yes	yes
Netherlands	US	24,037.58	34.48	yes	yes
China	US	22,055.15	36.18	no	yes
Israel	US	19,894.06	37.71	yes	yes
US	South Korea	19,740.63	39.23	yes	yes
US	Canada	19,688.13	40.75	yes	yes
South Korea	US	19,610.04	42.26	yes	yes
Australia	US	19,021.57	43.72	yes	yes
Italy	US	18,836.46	45.17	yes	yes
Sweden	US	18,631.36	46.60	yes	yes
US	France	16,180.48	47.85	yes	yes
France	Germany	3,035.6	49.02	yes	yes

Source: OECD Citations database from July 2014

The figures depicted in Table 2.5 for inventors' migration corridors have some points of similarity with those of citation corridors, although with slight but striking differences. Unsurprisingly, from this 20 country pair list the US appears as the most common host country for migrant inventors from 14 home countries. Migrant inventors from Chinese and Indian nationality to the US account for 24% or almost one fourth of all migrant inventors of our dataset. Besides, there is a high number of migrant inventors from Europe residing in the US mainly from the UK, Germany and France all of them being technology leading countries. When we focus on non-high income countries as sending countries there is more variety in migrant inventors' origin (see Table A3 in Appendix A), yet with the US as the main host country. Migrant inventors from China and India to the US altogether account for around 57%

of all migrant inventors from non-high income countries. Table A4 in Appendix A shows more than 77% of migrant inventors from low income countries live in high income countries – mainly the US, but also in Malaysia and China.

Table 2.5: Migration corridors (Total of inventor immigrants for the top 20 country pairs for the period 2006-2010)

Origin	Destination	Total migration	Cum. Migration share (%)	Origin high income	Destination high income
China	US	27,696	13.48	no	yes
India	US	21,712	24.04	no	yes
Canada	US	11,364	29.58	yes	yes
UK	US	8,313	33.62	yes	yes
Germany	US	5,895	36.49	yes	yes
Germany	Switzerland	4,952	38.9	yes	yes
South Korea	US	4,877	41.27	yes	yes
France	US	3,898	43.17	yes	yes
Japan	US	2,844	44.55	yes	yes
Russia	US	2,309	45.58	no	yes
France	Switzerland	1,880	46.59	yes	yes
Israel	US	1,878	47.51	yes	yes
Australia	US	1,783	48.37	yes	yes
Netherlands	US	1,670	49.19	yes	yes
France	Germany	1,492	49.91	yes	yes
Italy	US	1,492	50.64	yes	yes
China	Japan	1,463	51.35	no	yes
Germany	Netherlands	1,335	52	yes	yes
Austria	Germany	1,308	52.63	yes	yes
Turkey	USA	1,233	53.24	no	yes

Source: WIPO Statistics Database, October 2013

Trends of citations and inventors' migration in selected countries

We now turn to the comparison of both migrant inventors and knowledge flow trends for some of the biggest players as shown from the aforescribed migration and knowledge corridor tables.

In the following graphs, the years are put in the x-axis. The first y-axis reports the share of migrant inventors, which is the number of migrant inventors from one nationality residing in a given country over the total number of migrant inventors for a given year. The second y-axis reports the share of knowledge flow defined as the number of citations received by a given country from one country over the total citations of the given year in the whole sample. For an accurate view of the different fluctuations over the time we use different scales for representing some of the migration/knowledge graphs.

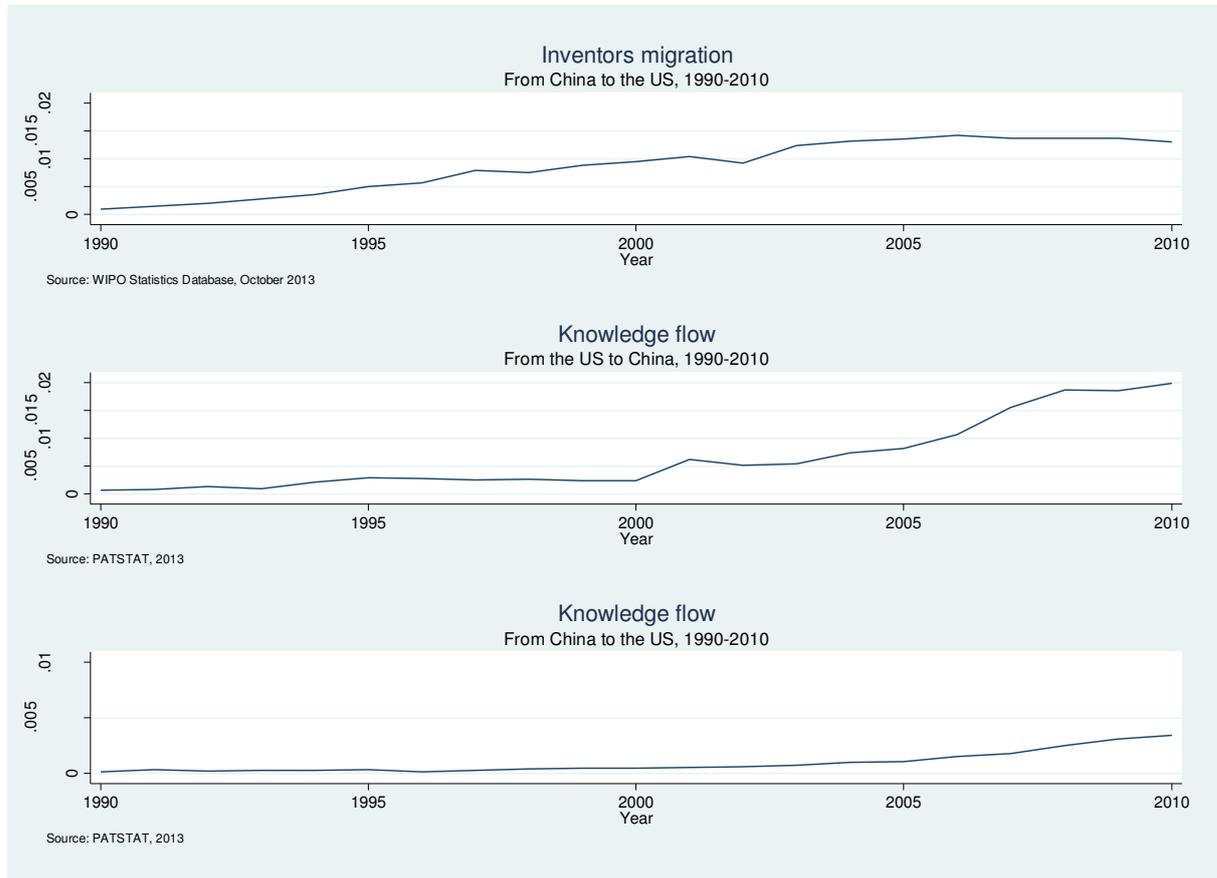
Figure 2.1 shows three curves for the studied period: the first one is the migration trend of Chinese migrant inventors in the US; the second one depicts knowledge flow trend from the US to China; and the last one represents knowledge flow trend from China to the US. All the three curves exhibit a general upward trend. The steady increase in the share of Chinese knowledge workers in the US seems to be in line with the steady increase in the share of citations received by US inventors, from Chinese ones on one hand; and the steady increase in the share of citations received by Chinese inventors from those from the US on the other hand.

From the first graph, we see the share of migrant inventors of Chinese nationality to the US has gradually increased from an initial value close to zero in 1990. However, there are two periods of slight decrease in this trend, namely the periods from 1997 to 1998 and right after 2001 to 2002 respectively. The first period corresponds to the years following the Illegal Immigration Reform and Immigrant Responsibility Act and the Welfare Bill passed in 1996. These are policies reinforcing the laws against illegal immigration and employment of undocumented immigrants in particular, and reducing access to public benefits for both documented and undocumented immigrants. The second period coincides with the post September 11th terrorist attacks and the increase of immigration laws enforcement during that period. However, the trend of citations does not seem to have been affected much by these two events. This implies immigration laws might be a good instrument for our present analysis.

The last two graphs shows similarly the rate of citations from Chinese inventors to US inventors, and the rate of citations from US inventors to Chinese. As can be seen, both have been increasing over the years. Admittedly, the former rate has been much higher than the latter throughout the years (which of course is partially driven by the much more advanced

technological level of the US). Yet, when comparing these two trends there seems to be some reciprocity in the knowledge exchange between China and the US.

Figure 2.1: Migration and knowledge trends comparison for China and the US



Similarly, a general upward trend is observed when looking at the graphs depicting trends in the rates of migrant inventors with Indian nationality in the US, citations from the US to India and citations from India to the US respectively (Figure 2.2). Additionally, here the migration curve slightly drops during the periods 1997-1998 and 2001-2002 too.

Figure 2.2: Migration and knowledge trends comparison for India and the US



Overall, the above comparison of the trends of knowledge flow and inventor's migration suggests there could be a relationship between both variables. However, from the reciprocity observed above, we assume migrant inventors might not only be impacting on the outflows of knowledge back to their sending economies from the receiving countries, but they might as well be a channel of knowledge to their receiving countries. An empirical analysis would allow us to gain more insight into the nature of these relationships and other determinants of knowledge flows.

2.4 Results

2.4.1 Skilled migration and brain gain

In the following regressions our key explanatory variable is the total number of active inventors of nationality i in country j in time t . We add a constant (1) to this variable in order to allow for its logarithmic transformation. This same logarithmic transformation has been applied to all the other control variables – except for dummy variables – used to explain the flow of citations from country i to country j . In this section we investigate the determinants of outflows of knowledge sent to the migrant inventors home country i from their host country j . We first present results from the baseline model regression, we then conduct some robustness check and we end with some instrumental variables (IV) results.

2.4.1.1 Baseline estimations

Table 2.6 reports the results from the baseline PPML regression, with robust country-pair clustered standard errors. First column returns results from the simple model specification without any interaction of control variables. The estimator of our main variable of interest, the migrant inventors' variable, has a value of 0.083, positive and statistically significant. This means that a 100% increase in the number of inventors from one source country in a destination country leads to 8.3 percentage point increase in the knowledge flow from that destination country to the source country. In our sample, the average value of country-pair migrant inventors is 1.011 and the average value of country-pair citations is 7.231 (see Appendix A, Table A5). Thus, 100% increase in migrant inventors (1.011 or about one migrant inventor) increases average citations or knowledge flow by 0.6 ($7.231 \times 8.3 \times 0.01$). Therefore, the estimated knowledge flow coefficient may appear small but it implies a significant increase in the amount of knowledge flow relative to a given increase in the amount of migrant inventors.

As for the coefficients for the countries contiguity and colonial ties variables, we don't find any significant result although they are positive. While cultural ties appear to be an important factor of knowledge flow as the estimator of the common language variable is positive and significant – 0.111. As expected, the coefficient for technological similarity is positive and

significant; with a value of 1.480 – close to what has been found by (MacGarvie, 2005) – confirming the hypothesis that countries that are close in their technological space are more likely to exchange knowledge. Another expected result is the negative and significant coefficient of the distance variable, meaning that knowledge flows fade with distance. We also find strong and positive estimated coefficients for the variables measuring each country inventive capacity – number of inventors. However, the coefficient for the inventive capacity is almost twice as large for sending or citing countries – 0.496 – as for receiving or cited countries – 0.282.

Next, we are interested to see how our variable of interest interaction with country pairs inherent ties, namely historical and cultural ties, would affect knowledge flows. We therefore add consecutively two new variables to our baseline model. Those variables represent respectively the product of the migration variable and colonial ties dummy – results shown in column (2) – and the product of the migration variable and common language dummy – results shown in column (3). Interestingly, we get negative signs and significance in the coefficients for both interaction variables in columns (2) and (3). This can be interpreted as follows: in the absence of historical and cultural ties between two countries, migration serves to ease knowledge exchanges between two countries. In other words, the importance of highly-skilled migrants as a bridge to knowledge flows between a pair of countries is much relevant when these countries do not share any common history or cultural background.

In order to get rid of the assumption that the results we get from the baseline estimation in column (1) might be attributed to knowledge flowing within company, we drop intra-company citations from our dependant variable. The new knowledge flow variable we get only captures inter-company citations. We do it by comparing company's names across citing and cited applicants and drop the citation each time we find a match between those names.²⁶ We thus run a new PPML regression with that dependant variable, whose results are shown in column (4). The major observation from these new results is that the coefficient for the migrant inventors' variable slightly drops to 0.080 and yet remains significant. This implies

²⁶ For the matching, we proceeded in 5 main steps. First we clean applicant names in the cited and citing patents by removing all punctuations and special characters and correcting for abbreviations. Then we compute the edit distance of the full applicant name using the Compged command in SAS, which is a command assigning a score based on how similar or different two strings are from each other. The closer strings are, the smaller that score will be. This first score helps us identify citing and cited applicants which are perfect match; that we drop. We thus proceed further by parsing applicant name strings into different columns and making a combination comparing citing and cited applicant strings. In a further step we manually check citing and cited applicants with at least two strings having a relatively low edit distance score and drop those we find being the same company, by additionally checking with their respective addresses.

there might be some international mobility of highly-skilled workers within MNCs (Multinational Corporations), who are the actual mechanisms of knowledge diffusion. However, this contribution is only a small part of a bigger phenomenon. Therefore, even though we will still be showing results for the regressions with this new dependent variable in the following tables, for the rest of analysis our main focus will be on our initial one – the one with both intra and inter-company citations.

Table 2.6: Knowledge outflows PPML baseline estimations for the period 1990 – 2010

	(1) All citations	(2) All citations	(3) All citations	(4) Without intra-company citations
ln(migrant inventors + 1)	0.083*** (0.010)	0.084*** (0.011)	0.089*** (0.011)	0.080*** (0.011)
Contiguity	0.022 (0.024)	0.020 (0.024)	0.034 (0.024)	0.008 (0.024)
Colony	0.020 (0.017)	0.129*** (0.037)	0.023 (0.017)	0.018 (0.017)
Common official lang.	0.111*** (0.019)	0.117*** (0.020)	0.241*** (0.028)	0.104*** (0.020)
ln(distance)	-0.092*** (0.012)	-0.090*** (0.012)	-0.089*** (0.012)	-0.090*** (0.012)
Colony*ln(migrants + 1)		-0.022*** (0.008)		
Common lang. *ln(migrants + 1)			-0.028*** (0.006)	
Technological similarity	1.480*** (0.076)	1.479*** (0.076)	1.495*** (0.076)	1.505*** (0.077)
ln(# of inv. in country i + 1)	0.496*** (0.030)	0.495*** (0.030)	0.495*** (0.030)	0.499*** (0.031)
ln(# of inv. in country j + 1)	0.282*** (0.030)	0.281*** (0.032)	0.282*** (0.032)	0.280*** (0.033)
Constant	-7.500*** (0.310)	-7.526*** (0.310)	-7.632*** (0.311)	-7.506*** (0.309)
Observations	399,443	399,443	399,443	399,443
Pseudo R2	0.982	0.982	0.982	0.981
Country iFE	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.1.2 Are the biggest players driving our results?

We replicate the PPML regressions for the baseline model removing from our data all observations with China and India as sending or citing countries and the US as a receiving or cited country consecutively. Again here we run the regressions for our main dependent variable and for the one where we remove intra-company citations, with country and time fixed effects. The results are displayed in Table 2.7. Overall except for some variations in their magnitude, those results are pretty similar to those from the baseline model. The first two columns, (1) and (2), show results of the regressions when we remove from the initial dataset all country pairs with the US as the receiving or cited country. In columns (3) and (4) we drop all pairs with India or China as the sending or citing country. The last two columns report the results of regressions when removing all country pairs with all of these three countries at the above mentioned positions. In all three cases, the coefficients for the migration variable remain positive and significant. The value of the coefficient decreases for the regressions without the US – 0.068 and 0.063 – and increases for the regressions from which we have dropped both India and China – 0.104 and 0.100, keeping in mind that the coefficient value in the baseline model was 0.083. The first two columns results confirm the key position of the US as an important source of knowledge and destination country for inventors, in so far that dropping that country produces a downward effect leading to a decrease of our baseline coefficient by one fourth. As for columns (3) and (4) the results suggest that including India and China in our dataset seem to lower down the coefficient for our focal variable. This implies that a variation in the knowledge flow to China and India is less sensible to a variation in the number of migrant inventors of Chinese and Indian origin abroad respectively. This can be attributed to the relatively high number of migrant inventors from China or India.

Table 2.7: Knowledge outflows PPML estimations without some of the biggest players for the period 1990 – 2010

	(1)	(2)	(3)	(4)	(5)	(6)
	All citations	Without intra- company citations	All citations	Without intra- company citations	All citations	Without intra- company citations
Ln(migrant inventors + 1)	0.068*** (0.012)	0.063*** (0.013)	0.104*** (0.011)	0.100*** (0.011)	0.079*** (0.012)	0.073*** (0.012)
Contiguity	0.039 (0.024)	0.023 (0.025)	0.012 (0.020)	-0.001 (0.021)	0.032 (0.021)	0.017 (0.022)
Colony	0.031 (0.023)	0.030 (0.024)	0.005 (0.017)	0.003 (0.018)	0.028 (0.023)	0.027 (0.024)
Common official lang.	0.101*** (0.022)	0.091*** (0.022)	0.114*** (0.018)	0.107*** (0.019)	0.100*** (0.021)	0.090*** (0.021)
Ln(distance)	-0.100*** (0.013)	-0.098*** (0.013)	-0.079*** (0.011)	-0.077*** (0.011)	-0.093*** (0.012)	-0.092*** (0.012)
Technological similarity	1.690*** (0.076)	1.718*** (0.077)	1.457*** (0.064)	1.480*** (0.066)	1.656*** (0.068)	1.684*** (0.070)
Ln(# of inv. in country i + 1)	0.483*** (0.034)	0.488*** (0.034)	0.356*** (0.031)	0.359*** (0.032)	0.334*** (0.036)	0.339*** (0.037)
Ln(# of inv. in country j + 1)	0.394*** (0.037)	0.395*** (0.037)	0.261*** (0.030)	0.258*** (0.031)	0.382*** (0.035)	0.381*** (0.036)
Constant	-7.734*** (0.363)	-7.731*** (0.363)	-7.631*** (0.299)	-7.642*** (0.298)	-7.838*** (0.355)	-7.837*** (0.355)
Observations	396,527	393,695	393,737	393,737	390,863	388,031
Pseudo R2	0.977	0.976	0.984	0.983	0.979	0.979
Country i FE	Yes	Yes	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Without the US as cited/receiving country	Yes	Yes	No	No	Yes	Yes
Without India as citing/ sending country	No	No	Yes	Yes	Yes	Yes
Without China as citing/sending country	No	No	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.1.3 Are there differences between income group pairs?

We now turn to an analysis where we focus on income group pairs. More precisely we are interested to see whether all income group countries get similar benefits in terms of knowledge flow from sending their highly-skilled to other countries on one hand and to high income countries on the other hand. We are particularly interested here on the case of low income sending or citing countries as it has been shown by empirical studies that they are the ones suffering the most from an intensive brain drain (Docquier et al., 2007; Docquier & Rapoport, 2012). For doing so, we build new explanatory variables by interacting the migration variable with dummies for each different income group pair. In columns (1) and (2) of Table 2.8, the first three explanatory variables correspond to the product of the migration variable and the dummies for respectively the pairs made of low, middle and high income as sending or citing and all countries as receiving or cited. The first three coefficients of columns (3) and (4) correspond to the explanatory variables which are the product of the migration variable and the dummies for respectively the pairs made of low, middle and high income as sending or citing and high income as receiving or cited. In general, the results show that for all income group pairs the relationship between highly-skilled migrants and knowledge flow remains positive and significant. However, there are striking differences that emerge from those results. Furthermore, when comparing columns (1) and (2) with columns (3) and (4), we see there is not much of a difference in the results of the coefficients for the explanatory variables with high income as receiving or cited and those with all countries as receiving or cited, no matter the income group of sending or citing countries. This implies that almost all of the effect of highly-skilled migration on citations shall be attributed to high income as receiving or cited countries. This is in line with our statistical description for the citations and inventors migration corridors as presented in the previous section, which has shown high income countries to be the top citations receiving and migrants' destination countries. Therefore we focus on columns (3) and (4) for the rest of our analysis.

One interesting result from these last two columns of Table 2.8 is the high gap between the coefficient for the migration variables for low-high income group pair – 0.270 – and the coefficient for the migration variable for the two other income group pairs – 0.045 and 0.086 respectively. This can be interpreted as a variation in the number of migrant inventors from low to high income countries would be likely to impact more on a variation in the knowledge flowing from high back to these low income countries, as compared with the impact induced

by other income groups. This does not necessarily mean for low income countries that impact will be bigger than for other income groups in absolute terms. For a better illustration, we use the means of our main variables for each income group. First, a 100% increase in the average of migrant inventors from low to high income countries which has a value of 0.097 in our sample – see Appendix A, Table A6 – would lead to a 27% increase in the average citations received by high from low income countries– 0.018. More precisely, if the number of migrant inventors from low income countries in high income ones increase by 0.097, then the outflows of knowledge to low income countries from the high income ones will increase by around $0.005 - 0.018 \times 27 \times 0.01$. By analogy, an increase in the number of migrant inventors from middle income countries in high income ones by 2.324 would lead to an increase in the knowledge flowing from high income to middle income countries by 0.115. While for high-high income group pairs, that impact corresponds to an increase of average migrant inventors by 5.850 leading to an increase of knowledge flow by 5.632. Putting it simply, if a low, middle and high income country supplies on average one additional migrant inventor to a high income country, the knowledge flow from that high to the low, middle and high income countries would increase by 0.052, 0.049 and 0.963 respectively.

Overall the results presented above are pretty optimistic, thus in favor of a brain gain effect from highly-skilled migration from all income groups to the high income ones. Even low income economies would benefit from sending their highly-skilled to high income economies. The elasticity of this effect is relatively higher to the impact on middle and high income economies, even though its value is much lower in absolute terms.

Table 2.8: Knowledge outflows income group PPML estimations for the period

	1990 - 2010			
	(1) All citations	(2) Without intra-company citations	(3) All citations	(4) Without intra-company citations
ln(migrant inv. + 1) Low income - All	0.279*** (0.069)	0.278*** (0.070)		
ln(migrant inv. + 1) Middle income - All	0.050*** (0.015)	0.047*** (0.015)		
ln(migrant inv. + 1) High income - All	0.092*** (0.011)	0.088*** (0.011)		
ln(migrant inv. + 1) Low - High income			0.270*** (0.069)	0.268*** (0.072)
ln(migrant inv. + 1) Middle - High income			0.045*** (0.015)	0.042*** (0.015)
ln(migrant inv. + 1) High - High income			0.086*** (0.011)	0.083*** (0.011)
Contiguity	0.018 (0.024)	0.004 (0.025)	0.017 (0.024)	0.003 (0.025)
Colony	0.013 (0.017)	0.011 (0.017)	0.012 (0.017)	0.010 (0.018)
Common official lang.	0.110*** (0.019)	0.102*** (0.020)	0.115*** (0.020)	0.107*** (0.020)
ln(distance)	-0.089*** (0.012)	-0.088*** (0.012)	-0.093*** (0.012)	-0.091*** (0.012)
Technological similarity	1.475*** (0.074)	1.500*** (0.075)	1.480*** (0.074)	1.504*** (0.075)
ln(# of inv. in country i + 1)	0.502*** (0.030)	0.505*** (0.031)	0.503*** (0.030)	0.506*** (0.031)
ln(# of inv. in country j + 1)	0.279*** (0.032)	0.278*** (0.033)	0.289*** (0.032)	0.287*** (0.033)
Constant	-7.629*** (0.303)	-7.637*** (0.302)	-7.603*** (0.303)	-7.614*** (0.302)
Observations	399,443	399,443	399,443	399,443
Pseudo R2	0.982	0.981	0.981	0.981
Country I FE	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.1.4 Instrumental variable strategy

In order to correct for potential endogeneity and omitted variables bias to our model caused by unobserved factors likely to explain knowledge flow between two countries, we run an instrumental variable (IV) test using two instruments separately: an index of immigration policies and the 1960 stock of immigrants in a given country. We assume each of these two variables to be good instruments in so far that they are likely to explain highly-skilled migration without being correlated with patent citations or knowledge flow

The first instrument comes from Ortega & Peri (2013). It is computed yearly for 12 OECD countries as host countries and all other countries of the world as home countries, for the period 1980-2006. Each of the host countries is initialized to zero in the first year. If there is no relevant immigration policy occurring, the variable remains constant. When in a year an immigration policy tightening up entry conditions is passed, the variable increases its value by one. While for the relaxation of entry condition, the variable decreases the degree of tightness by one. We lag this index by three years in order to account for the time needed for immigration laws to be assimilated and effectively start impacting on the flow of migrants. The availability of the index of immigration laws for only 12 countries considerably reduces our observations down to 31,024 in the second panel. The second instrument is the 1960 stock of immigrants from one origin country in a host country (Docquier et al., 2009). Unlike the immigration law variable which is time variant, this variable remains constant throughout the period covered by our sample for each country pair.

In order to be certain about the validity of our instruments, we run a first-stage OLS regression whose results are shown in Table 2.9. Columns (1) and (2) report regressions respectively for the 1960 stock of immigrants and immigration policies as explanatory variables to migrant inventors. Both columns show our two instruments to be strongly correlated with our suspected endogenous variable. Also, the coefficients for each of the instruments have the expected signs; there is a positive relationship between the 1960 immigrants stock in a given country and future inflows of highly-skilled migrants to that country. We find a negative relationship between restrictive immigration laws in a given country and future inflows of highly-skilled migrants to that country.

Therefore, we run IV poisson regressions with each of these two variables as an instrument, this for our initial dependent variable – columns (1) and (3) of Table 2.10 – and for the

dependent variable without intra-company citations – columns (2) and (4) of Table 2.10. The results depicted in first row of Table 2.10 confirm the positive impact of migrant inventors on knowledge flowing back to their sending countries from the host ones. We find the coefficients for the migrant inventors’ variable to be of an order of 0.258, 0.239, 0.260, 0.268 in columns (1), (2), (3) and (4) respectively. Thus we still get positive and statistically significant results even after having instrumented our dependent variable.

Table 2.9: First stage OLS: determinants of inventors’ migration

	(1) Migrant inventors	(2) Migrant inventors
ln(immigrants60)	0.021*** (0.001)	
Immigration policies 3-years lagged		-0.010** (0.004)
Contiguity	0.115*** (0.010)	1.019*** (0.054)
Colony	0.063*** (0.013)	0.322*** (0.027)
Common official lang.	-0.001 (0.003)	0.138*** (0.016)
ln(distance)	-0.032*** (0.001)	-0.118*** (0.011)
ln(# of inv. in country i + 1)	0.018*** (0.001)	0.098*** (0.006)
ln(# of inv. in country j + 1)	0.015*** (0.001)	-0.088*** (0.026)
Constant	0.149*** (0.013)	1.239*** (0.181)
Observations	393,674	31,024
R2	0.347	0.624
Country I FE	Yes	Yes
Country j FE	Yes	Yes
Year FE	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

Table 2.10: IV poisson regressions for knowledge outflows

	(Immigrants 60) (1) All citations	(Immigrants 60) (2) Without intra- company citations	(Immigration policies) (3) All citations	(Immigration policies) (4) Without intra- company citations
Ln(migrant inventors + 1)	0.258*** (0.070)	0.239*** (0.069)	0.260** (0.103)	0.268** (0.103)
Contiguity	0.013 (0.026)	0.001 (0.026)	-0.001 (0.024)	-0.015 (0.024)
Colony	0.019 (0.019)	0.017 (0.019)	-0.022 (0.020)	-0.024 (0.020)
Common official lang.	0.011 (0.047)	0.011 (0.048)	0.061 (0.058)	0.048 (0.059)
ln(distance)	0.011 (0.037)	0.003 (0.037)	0.025 (0.062)	0.0325 (0.062)
Technological similarity	1.307*** (0.108)	1.345*** (0.109)	1.128*** (0.116)	1.144*** (0.117)
ln(# of inv. in country i + 1)	0.457*** (0.033)	0.464*** (0.034)	0.443*** (0.041)	0.442*** (0.041)
ln(# of inv. in country j + 1)	0.260*** (0.031)	0.261*** (0.032)	-0.023 (0.039)	-0.036 (0.040)
Constant	-8.060*** (0.366)	-8.016*** (0.364)	0.178 (0.865)	0.207 (0.867)
Observations	393,674	393,674	40,530	40,530
Pseudo R2	0.979	0.980	0.981	0.980
Country I FE	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.2 Immigration and innovation

In section 2.4.1, our main investigation was centered on the outcomes from highly-skilled migration in terms of the outflows of knowledge from the receiving to the sending countries. In other words, the direction of knowledge was from the receiving–country j – to the sending country – country i . We now turn to an analysis of the impact of highly-skilled migration on the receiving countries themselves. Basically, we look for whether there are some knowledge inflows to the receiving country – country j or citing country – brought by migrant inventors from their home country – country i or cited country. In the following, we first present results from the baseline model then conduct some robustness check and end with some IV strategy.

2.4.2.1 Baseline estimations for knowledge inflows

In Table 2.11 we see the results from the baseline estimations with our initial citations variable – column (1) – and the citations variable without intra-company citations – column (2). What interests us the most here is the first row which shows results for the coefficients of the migrant inventors’ variable, since other results are not far from what we got in the knowledge outflows baseline model. In the first column we get a coefficient of 0.060, which is lower from the 0.083 value we got in the baseline model of the knowledge outflows section. That is, a 100% increase in the total migrant inventors from one sending country to a destination would lead to a 6% increase in the knowledge flowing from the sending country to that destination country. In our dataset, on average we have 7.50 citations from citing-cited or receiving-sending country pairs and the average migrant inventors for these country pairs is 1.048 – Appendix A, Table A6. Thus a 100% increase in migrant inventors would mean an additional 1.048 migrants, which will produce an increase of citations by 0.45 – $7.50 \times 6 \times 0.01$ –from sending to receiving countries. In other words, on average one additional migrant inventor sent to a host country will increase the knowledge flowing to that host country from the home country by 0.43. This value is slightly smaller than the 0.6 increase we find in the baseline model for the knowledge outflows. In column (2) the coefficient for the migration variable has a value of 0.056, which is just slightly lower than the coefficient we got in the first column. Therefore, similarly to the knowledge outflows there seems to be a small part of the mechanism of knowledge brought to destination countries by MNCs, in which migrant inventors are employees.

Table 2.11: Knowledge inflows from migrant inventors 1990 – 2010

	PPML (1) All citations	PPML (2) Without intra-company citations
ln(migrant inventors + 1)	0.060*** (0.010)	0.056*** (0.010)
Contiguity	0.007 (0.024)	-0.006 (0.024)
Colony	0.028 (0.017)	0.025 (0.017)
Common official lang.	0.131*** (0.019)	0.123*** (0.019)
ln(distance)	-0.115*** (0.012)	-0.113*** (0.012)
Technological similarity	1.494*** (0.075)	1.520*** (0.076)
ln(# of inv. in country i + 1)	0.284*** (0.033)	0.282*** (0.034)
ln(# of inv. in country j + 1)	0.501*** (0.030)	0.504*** (0.030)
Constant	-7.364*** (0.309)	-7.374*** (0.308)
Observations	385,094	385,094
Pseudo R2	0.981	0.980
Country I FE	Yes	Yes
Country j FE	Yes	Yes
Year FE	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.2.2 Are the biggest players driving our results?

Again, we separately drop the US, India and China to check whether these three countries are driving our results for the knowledge inflows. Results from our regressions are shown in Table 2.12. The main remark coming from these results is that the coefficients for our migration variable in all the panels do not differ much from the coefficient we find in baseline model; which was not the case for the knowledge outflows results in the previous section. This suggests these three countries might only weight more when it comes to the mechanism of knowledge outflows from highly-skilled migration than for the one of knowledge inflows.

Table 2.12: Knowledge inflows PPML estimations without some of the biggest players for the period 1990 – 2010

	(1)	(2)	(3)	(4)	(5)	(6)
	All citations	Without intra- company citations	All citations	Without intra- company citations	All citations	Without intra- company citations
ln(migrant inventors + 1)	0.059*** (0.008)	0.054*** (0.009)	0.063*** (0.010)	0.059*** (0.010)	0.059*** (0.008)	0.054*** (0.008)
Contiguity	0.005 (0.023)	-0.010 (0.024)	0.001 (0.023)	-0.013 (0.023)	0.001 (0.022)	-0.014 (0.023)
Colony	0.014 (0.018)	0.012 (0.018)	0.023 (0.017)	0.021 (0.017)	0.011 (0.018)	0.010 (0.018)
Common official lang.	0.181*** (0.020)	0.172*** (0.020)	0.136*** (0.018)	0.128*** (0.019)	0.186*** (0.019)	0.177*** (0.019)
ln(distance)	-0.101*** (0.012)	-0.097*** (0.012)	-0.113*** (0.012)	-0.111*** (0.012)	-0.098*** (0.012)	-0.095*** (0.012)
Technological similarity	1.448*** (0.075)	1.478*** (0.077)	1.511*** (0.076)	1.537*** (0.077)	1.462*** (0.077)	1.494*** (0.079)
ln(# of inv. in country i + 1)	0.197*** (0.029)	0.196*** (0.029)	0.223*** (0.036)	0.221*** (0.037)	0.130*** (0.031)	0.128*** (0.032)
ln(# of inv. in country j + 1)	0.523*** (0.035)	0.526*** (0.036)	0.498*** (0.030)	0.502*** (0.030)	0.525*** (0.035)	0.528*** (0.035)
Constant	-7.547*** (0.337)	-7.564*** (0.336)	-7.447*** (0.314)	-7.456*** (0.314)	-7.652*** (0.343)	-7.668*** (0.342)
Observations	376,647	376,647	379,438	379,430	371,025	371,018
Pseudo R2	0.982	0.982	0.981	0.981	0.985	0.982
Country I FE	Yes	Yes	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Without the U.S. as cited/receiving country	Yes	Yes	No	No	Yes	Yes
Without India as citing/ sending country	No	No	Yes	Yes	Yes	Yes
Without China as citing/sending country	No	No	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.2.3 Are there any differences between income group pairs?

Next we check if there is any difference in the relationship between migrant inventors and citations for the different income group pairs. We replicate the same exercise as in the previous section, adapting it to the knowledge inflows logic. That is, we use the migrant inventor variables for income group pairs we had built previously and regress them to citations from citing as the receiving to cited as the sending countries. As explained above, we build income group pair migration variables by multiplying our country pair migration variable by dummies for each income group pair. There are two types of income group pairs: low, middle and high to all other countries, and low, middle and high to high income countries. We thus regress our citation variables – all citations and the variable without intra-company citations – on the migration variables for each of these two types of income group pairs separately. The results from these regressions are presented in Table 2.13. Again here we only comment on the last two columns showing results for the low, middle and high to high income group pairs, since high income countries seem to be the ones driving the most of our results. One interesting remark that pops out from all these regressions for knowledge inflows is the coefficient for middle to high income group pair which has lost its significance. This suggests migrant inventors from middle income countries living in high income ones do not seem to convey knowledge inflows to their high income host countries from their middle income countries. Our data shows most of the middle income effect is accounted for by China and India. Therefore, our result for the middle to high income group pair is not surprising since we have seen in the previous section that dropping these two countries do not affect much our result for the knowledge inflows baseline model. The last two columns show the coefficients for the migrant inventors' variable for low to high income group pair – 0.162 and 0.173 – to be higher than for other income group pairs and significant. For our initial “all citations” variable – column (3) – we find a 100% increase in the total migrant inventors from low to high income countries would lead to an increase of the total citations from high income received by low income countries by 16.2%. While a similar increase in the number of migrant inventors from high to high income countries would lead to a 4.9% increase in the citations from high to high income countries. A quick exercise similar to the one we have conducted for the knowledge outflows shows this higher value of the coefficients for low to high income groups does not translate to what we find in absolute terms. Indeed, on average there is 0.022 citations from high to low income countries and 0.155 migrant inventors from

low income countries in the high income ones. Therefore, an increase of migrant inventors from low to high income countries by 0.097 would lead to an increase of citations from high to low income by $0.002 - 0.014 \times 16.2 \times 0.01$. On the other hands, there is in our dataset an average of 65.487 citations from high to high income countries and an average of 5.850 migrant inventors from high income countries in the high income ones. An increase in the number of migrant inventors from high to high income countries by 5.850 would lead to an increase in the number of citations from high to high income groups by 3.209. Overall, this means on average one additional inventor migrating from a high to another high income country would lead to an increase of citations from the latter to the former by 0.548. While for the low to high income group pair, one additional migrant inventor would lead to an increase of citations from the high to the low income country by 0.021.

Table 2.13: Knowledge inflows income group PPML estimations for the period 1990 - 2010

	(1) All citations	(2) Without intra-company citations	(3) All citations	(4) Without intra-company citations
ln(migrant inventors + 1) Low income– All	0.176** (0.073)	0.187** (0.077)		
ln(migrant inventors + 1) Middle income - All	0.015 (0.016)	0.008 (0.016)		
ln(migrant inventors + 1) High income - All	0.062*** (0.010)	0.058*** (0.010)		
ln(migrant inventors + 1) Low - High income			0.162** (0.073)	0.173** (0.077)
ln(migrant inventors + 1) Middle - High income			0.002 (0.016)	-0.004 (0.016)
ln(migrant inventors + 1) High - High income			0.049*** (0.009)	0.045*** (0.010)
Contiguity	0.004 (0.024)	-0.010 (0.024)	0.008 (0.024)	-0.006 (0.025)
Colony	0.026 (0.017)	0.023 (0.017)	0.020 (0.017)	0.018 (0.017)
Common official lang.	0.132*** (0.019)	0.124*** (0.019)	0.141*** (0.019)	0.132*** (0.019)
ln(distance)	-0.115*** (0.012)	-0.113*** (0.012)	-0.120*** (0.012)	-0.118*** (0.012)
Technological similarity	1.491*** (0.075)	1.517*** (0.076)	1.507*** (0.074)	1.532*** (0.075)
ln(# of inv. in country i + 1)	0.287*** (0.033)	0.285*** (0.034)	0.287*** (0.033)	0.286*** (0.034)
ln(# of inv. in country j + 1)	0.500*** (0.030)	0.503*** (0.030)	0.518*** (0.030)	0.520*** (0.030)
Constant	-7.421*** (0.314)	-7.439*** (0.314)	-7.409*** (0.314)	-7.427*** (0.314)
Observations	385,094	385,094	385,094	385,094
Pseudo R2	0.981	0.980	0.980	0.980
Country i FE	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1

2.4.2.4 Instrumental variable strategy for the inflows of knowledge

For the knowledge inflows we run some IV regressions, but this time using only the 1960 stock of immigrants as the immigration policies instrument would not be a suitable instrument here. This is to the extent that policy makers can modify migration policies precisely to attract knowledge flows while it is less likely to have policy makers purposely changing migration policies in order to modulate knowledge outflows. This can be explained by the fact that immigration policies are more likely to be correlated with knowledge inflows, which was not the case for knowledge outflows.

The first-stage OLS regression results are shown in Table 2.9 above. We run IV poisson regressions for our initial ‘all citations variable’ and the dependant variable ‘without intra-company citations’, whose results are shown in Table 2.14. Our main result is the positive sign and significance of the coefficient for the migrant inventors variable in both columns. Another interesting finding is that the coefficient of the variable for ‘Colony’ becomes statistically significant. This suggests past colonial ties seem to matter for having some knowledge inflow from highly-skilled migrants to their host countries. This was not the case for knowledge outflows home countries.

Table 2.14: IV poisson regressions for knowledge inflows

	(Immigrants 60) (1) All citations	(Immigrants 60) (2) Without intra-company citations
ln(migrant inventors + 1)	0.128*** (0.047)	0.110** (0.048)
Contiguity	-0.012 (0.029)	-0.022 (0.030)
Colony	0.036** (0.018)	0.032* (0.019)
Common official lang.	0.099*** (0.030)	0.097*** (0.030)
ln(distance)	-0.084*** (0.022)	-0.089*** (0.022)
Technological similarity	1.411*** (0.083)	1.453*** (0.085)
ln(# of inv. in country i + 1)	0.275*** (0.033)	0.275*** (0.034)
ln(# of inv. in country j + 1)	0.485*** (0.029)	0.492*** (0.034)
Constant	-7.499*** (0.222)	-7.479*** (0.313)
Observations	379,438	379,430
Pseudo R2	0.980	0.980
Country I FE	Yes	Yes
Country j FE	Yes	Yes
Year FE	Yes	Yes

Notes:*** p<0.01, ** p<0.05, * p<0.1

2.5 Discussion and conclusions

In this paper we have used the gravity model to show how highly-skilled migration— as measured by the number of migrant inventors – affects international knowledge diffusion – as measured by patent citations. Our results suggest migrant inventors living in a given country are an important channel of knowledge flow not only to their home countries but also to their host countries. This implies there are some knowledge externalities from the network developed around the ties that these migrants keep with their people home. Our estimates for the knowledge outflows – to the home countries – and for the knowledge inflows – to the host countries – remain positive and significant even after having dropped India, China and the US from the regressions, confirming a general brain gain for all countries from highly-skilled migrants and not only for the case of Chinese and Indian migrants in the US, as shown by previous studies. Contrary to what have been advocated in the migration literature on the detrimental effect of highly-skilled migration from low income countries and an alarming brain drain, we find that low income countries technologically benefit from their migrant inventors living in high income countries. While at the same time, highly-skilled migrants from low income countries also bring in some knowledge to their high income host countries from their low income sending ones. Interestingly, the absorptive capacity or the total number of highly-skilled within each country matters for these knowledge flows to be observed. This absorptive capacity is even more important for sending countries than for the receiving ones as we find the coefficient for the total number of inventors in the sending country to be higher than the one for the total number of inventors in the receiving country. This has some implications in terms of the type of policies on which should emphasize for countries that are the biggest suppliers of highly-skilled migrants, particularly the net supplier's ones. Those policies should be oriented towards investing in more education or technology oriented programs that would allow them to increase their internal highly-skilled capacity and so their inventive or absorptive capacity. Additionally, as a more general recommendation for all countries, our research intends to convey the message that, instead of focusing the debate on brain drain issues, the attention of home and host countries policy makers should be more oriented towards finding strategies in order to establish and strengthen connections between highly-skilled diaspora and those remaining home, through adequate knowledge networks. This is relevant to the extent that we find highly-skilled migration could be beneficial for both sending and receiving countries and this for all country groups.

Appendix A – Some additional tables

Table A1: Citations corridors – without high income as citing countries (Total flow of citations for the top 20 country pairs for the period 2006-2010)

Citing country	Cited country	Total share of citations	Citation Cum. (%)	Share	High income citing	High income cited
China	USA	22,055.15		20.07	no	yes
China	Japan	8,533.97		27.83	no	yes
India	USA	7,673.99		34.81	no	yes
China	South Korea	5,111.61		39.46	no	yes
China	Germany	4,143.47		43.23	no	yes
Russian Federation	USA	3,281.55		46.22	no	yes
South Africa	USA	2,885.64		48.85	no	yes
Brazil	USA	2,874.61		51.47	no	yes
China	UK	2,078.99		53.36	no	yes
India	Japan	1,747.44		54.95	no	yes
Mexico	USA	1,737.97		56.53	no	yes
China	France	1,654.54		58.04	no	yes
China	Finland	1,594.19		59.49	no	yes
Turkey	USA	1,451.02		60.81	no	yes
India	Germany	1,392.88		62.08	no	yes
China	Canada	1,384.37		63.34	no	yes
Malaysia	USA	1,193.37		64.43	no	yes
China	Sweden	1,162.83		65.49	no	yes
Turkey	Germany	1,019.76		66.42	no	yes
India	UK	965.89		67.3	no	yes

Source: WIPO Statistics Database, October 2013

Table A2: Citations corridors – with low income only as citing countries (Total flow of citations for the top 20 country pairs for the period 2006-2010)

Citing country	Cited country	Total share of citations	Citation Cum. (%)	Share	High income citing	High income cited
Kenya	USA	33.85		18.06	no	yes
North Korea	South Korea	10.04		23.42	no	yes
North Korea	USA	9.70		28.6	no	yes
North Korea	Germany	8.3		33.03	no	yes
Tajikistan	Germany	7.53		37.05	no	yes
North Korea	Japan	6.83		40.69	no	yes
Kenya	UK	5.59		43.67	no	yes
Sierra Leone	USA	4.33		45.98	no	yes
Kenya	Germany	4.24		48.24	no	yes
Kenya	Canada	4.08		50.42	no	yes
Burundi	USA	4		52.55	no	yes
Kenya	China	3.51		54.42	no	no
Madagascar	Germany	3		56.02	no	yes
North Korea	China	3		57.62	no	no
Burkina Faso	USA	3		59.22	no	yes
Zimbabwe	USA	3		60.82	no	yes
Kenya	Sweden	3		62.42	no	yes
Bangladesh	USA	2.99		64.02	no	yes
Uganda	USA	2.69		65.46	no	yes
North Korea	Austria	2		66.53	no	yes

Source: WIPO Statistics Database, October 2013

Table A3: Migration corridors without high income origin countries (Total of inventor immigrants for the top 20 country pairs for the period 2006-2010)

Origin country	Destination country	Total migration	Cum. Migration share (%)	High income origin	High income destination
China	USA	27,696	31.79	no	yes
India	USA	21,712	56.71	no	yes
Russian Federation	USA	2,309	59.36	no	yes
China	Japan	1,463	61.04	no	yes
Turkey	USA	1,233	62.46	no	yes
China	Singapore	1,149	63.78	no	no
Iran	USA	960	64.88	no	yes
Brazil	USA	763	65.75	no	yes
Mexico	USA	722	66.58	no	yes
Romania	USA	710	67.39	no	yes
Russian Federation	Germany	702	68.2	no	yes
India	Singapore	610	68.9	no	no
Malaysia	Singapore	607	69.6	no	no
Ukraine	USA	601	70.29	no	yes
China	Germany	555	70.92	no	yes
China	UK	545	71.55	no	yes
Malaysia	USA	484	72.11	no	yes
Argentina	USA	478	72.65	no	yes
South Africa	USA	414	73.13	no	yes
India	UK	393	73.58	no	yes

Source: WIPO Statistics Database, October 2013

Table A4: Migration corridors – with low income only as origin (Total of inventor immigrants for the top 20 country pairs for the period 2006-2010)

Origin country	Destination country	Total migration	Cum. Migration share (%)	High income origin	High income destination
Bangladesh	USA	188	16.97	no	yes
Nepal	USA	105	26.45	no	yes
Kenya	USA	92	34.75	no	yes
Ethiopia	USA	86	42.51	no	yes
North Korea	USA	59	47.83	no	yes
Bangladesh	Japan	54	52.7	no	yes
Tanzania	USA	40	56.31	no	yes
Uganda	USA	31	59.11	no	yes
Bangladesh	Canada	26	61.46	no	yes
Bangladesh	UK	22	63.45	no	yes
Zimbabwe	USA	21	65.35	no	yes
Ethiopia	Denmark	21	67.25	no	yes
Haiti	USA	17	68.78	no	yes
Bangladesh	Malaysia	15	70.13	no	no
Nepal	China	15	71.48	no	no
Niger	USA	14	72.74	no	yes
Nepal	Japan	14	74	no	yes
Bangladesh	Singapore	14	75.26	no	yes
Bangladesh	Finland	13	76.43	no	yes
Bangladesh	South Korea	13	77.6	no	yes

Source: WIPO Statistics Database, October 2013

Table A5: Descriptive statistics

Variables	Observations	Mean	Std. Dev.	Min	Max
ALL INCOME GROUPS					
Knowledge outflows					
All country-pair citations	399,443	7.231	171.279	0	20,888.45
Country-pair without inter-company citations	399,443	6.996	166.172	0	20,132.34
Migrant inventors	399,443	1.011	34.906	0	5,837
Knowledge inflows					
All country-pair citations	385,094	7.500	174.433	0	20,888.45
Country-pair inter-company citations	385,094	7.256	169.233	0	20,132.34
Migrant inventors	385,094	1.048	35.550	0	5,837
FROM LOW TO HIGH INCOME COUNTRIES					
Knowledge outflows					
All country-pair citations	18,683	0.018	0.233	0	12.167
Country-pair without inter-company citations	18,683	0.017	0.225	0	12.167
Migrant inventors	18,683	0.097	1.294	0	68
Knowledge inflows					
All country-pair citations	18,683	0.014	0.126	0	4
Country-pair without inter-company citations	18,683	0.013	0.122	0	4
Migrant inventors	18,683	0.097	1.294	0	68
FROM MIDDLE TO HIGH INCOME COUNTRIES					
Knowledge outflows					
All country-pair citations	65,894	2.545	51.352	0	5,694.547
Country-pair without inter-company citations	65,894	2.472	49.608	0	5,462.102
Migrant inventors	65,894	2.325	75.130	0	5,837
Knowledge inflows					
All country-pair citations	65,894	0.650	8.891	0	980.823
Country-pair without inter-company citations	65,894	0.616	8.303	0	895.514
Migrant inventors	65,894	2.325	75.130	0	5,837
FROM HIGH TO HIGH INCOME COUNTRIES					
Knowledge outflows					
All country-pair citations	40,812	65.487	528.165	0	20,888.45
Country-pair without inter-company citations	40,812	63.359	512.476	0	20,132.34
Migrant inventors	40,812	5.850	52.609	0	2,415
Knowledge inflows					
All country-pair citations	40,812	65.487	528.165	0	20,888.45
Country-pair without inter-company citations	40,812	63.359	512.476	0	2,415
Migrant inventors	40,812	5.850	52.609	0	2,415

Table A6: PPML regressions with explanatory variables time lagged for the knowledge outflows

	(1) All citations	(2) All citations	(3) Without intra-company citations	(4) Without intra-company citations
$[\text{Ln}(\text{migrant inventors} + 1)]_{t-1}$	0.068*** (0.012)		0.066*** (0.012)	
$[\text{Ln}(\text{migrant inventors} + 1)]_{t-2}$		0.056*** (0.007)		0.052*** (0.007)
Contiguity	0.034 (0.021)	0.028* (0.017)	0.020 (0.022)	0.014 (0.017)
Colony	0.016 (0.016)	0.201 (0.014)	0.014 (0.016)	0.018 (0.014)
Common official language	0.226*** (0.019)	0.129*** (0.015)	0.223*** (0.019)	0.121*** (0.015)
$\text{Ln}(\text{distance})$	-0.106*** (0.012)	-0.105*** (0.009)	-0.105*** (0.012)	-0.104*** (0.010)
$[\text{Technological similarity}]_{t-1}$	0.033*** (0.011)		0.031*** (0.011)	
$[\text{Technological similarity}]_{t-2}$		1.655*** (0.063)		1.679*** (0.064)
$[\text{Ln}(\# \text{ of inventors in country } i + 1)]_{t-1}$	0.693*** (0.023)		0.696*** (0.023)	
$[\text{Ln}(\# \text{ of inventors in country } i + 1)]_{t-2}$		0.732*** (0.019)		0.736*** (0.019)
$[\text{Ln}(\# \text{ of inventors in country } j + 1)]_{t-1}$	0.561*** (0.030)		0.559*** (0.031)	
$[\text{Ln}(\# \text{ of inventors in country } j + 1)]_{t-2}$		0.684*** (0.028)		0.684*** (0.028)
Constant	-9.254*** (0.280)	-8.852*** (0.283)	-9.273*** (0.280)	-8.867*** (0.283)
N	394,522	389,601	394,522	389,601
Pseudo R2	0.982	0.986	0.982	0.987
Country I FE	Yes	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Table A7: More explanatory variables for the knowledge outflows model

	PPML (1)	PPML (3)
	All citations	Without intra-company citations
Ln(migrant inventors + 1)	0.054*** (0.010)	0.050*** (0.010)
Contiguity	-0.011 (0.020)	-0.023 (0.021)
Colony	0.041*** (0.015)	0.038** (0.016)
Common official language	0.106*** (0.017)	0.099*** (0.017)
Ln(distance)	-0.032** (0.015)	-0.038*** (0.015)
Technological similarity	1.484*** (0.068)	1.512*** (0.069)
Ln(bilateral trade i/j)	0.088*** (0.012)	0.082*** (0.012)
Ln (GDP in country i)	0.518*** (0.108)	0.506*** (0.110)
Ln (GDP in country j)	2.168*** (0.127)	2.216*** (0.129)
Ln(# of inventors in country i + 1)	0.455*** (0.030)	0.461*** (0.031)
Ln(# of inventors in country j + 1)	0.138*** (0.027)	0.133*** (0.028)
Constant	-35.19*** (1.427)	-35.81*** (1.516)
N	252,674	252,674
Pseudo R2	0.985	0.984
Country I FE	Yes	Yes
Country j FE	Yes	Yes
Year FE	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Chapter 3

Highly-skilled migration and the internationalization of knowledge

3.1 Introduction

A growing literature has been dealing with the role of highly-skilled – hs – international migrants as a channel of knowledge exchange and circulation across countries and regions. This literature has exploited the underlying idea that sharing a common social and cultural background could favour different types of exchanges within specific diaspora groups. In particular, a common social and cultural background could support the formation and maintenance of social networks of hs migrants, where knowledge would be exchanged or circulate more easily, both within the migrants' destination countries and to their countries of origin. This assumption rests on the principle that scientific and technical knowledge contains tacit elements, whose transfer demand direct human interaction and some form of proximity (geographical, cultural...) (Breschi & Lissoni, 2009; Dosi, 1988; Jaffe & Caballero, 1993; Leonard & Sensiper, 1998). Some studies have shown that hs migrants from specific origin countries – most notably Indian and Chinese – form a strongly connected diaspora insofar that they tend to have a higher propensity to pass on knowledge to other hs migrants of same origin at the destination than with nationals (Agrawal et al., 2006; Breschi et al., 2015; Kerr, 2009). These findings highlight the importance of interactions or links among hs diaspora members of same origin. The set of these interactions form what we call hs diaspora knowledge networks. It is within such networks that part of new ideas and innovations are created within destination countries, thus contributing to their economic growth (Ackers, 2005; Agrawal et al., 2011; Gill, 2005; Kerr, 2008).

We observe, however, that there is no reason to presume that social interactions between same-origin migrants ought to be bound to the countries of destination or within the origin-destination axis. In fact, hs diaspora members might have a higher propensity to collaborate wherever they are, including across multiple destination countries. In other words, hs

diaspora knowledge networks may span far beyond one single destination country to two or more destinations. There have been some qualitative studies on the functioning of hs diaspora knowledge networks – particularly from developing countries – across destination countries depicting how they come together in an associative platform in order to channel knowledge back home (Adepoju et al., 2008; Brown, 2002; Meyer & Wattiaux, 2006). But, to the best of our knowledge, no empirical study has yet investigated how hs diasporas contribute to the internationalization of knowledge across destination countries. Moreover, little research has been done on the mechanisms underpinning knowledge transfer among destination countries within immigrant networks.

We thus intend to fill the existing gap in the literature by assessing the impact on collaboration in innovative/knowledge activities of two large hs diasporas – Chinese and Indian. We refer to Science & Technology (S&T) collaboration for a large sample of the Organisation for Economic Co-operation and Development (OECD) destination countries. In particular, we explore two variables as proxies for S&T collaboration: co-inventorship and co-authorship, as computed from various data sources. For each of these dependent variables, we apply a gravity approach at a country pair level. We run PPML – Poisson Pseudo Maximum Likelihood – regressions (Santos-Silva & Tenreyro, 2006) for each of the two variables. Our preliminary results suggest a positive impact of these hs diasporas on all our knowledge collaboration variables. Additionally, we test for the impact of other hs diasporas, all from top hs migrant sending countries to OECD host countries, and find positive results for Vietnam, Pakistan and Iran. Although the number of Indian and Chinese hs diasporas are a way larger, these other hs diasporas produce similar effects on co-inventorship and co-authorship.

The rest of the paper is organized as followed: in section 2 we present a review of the literature, in section 3 we briefly discuss key definitions, while in section 4 we develop our methodology. In section 5 we discuss our results, and finally in the last section we conclude.

3.2 Literature review

3.2.1 Migration, social networks and innovation

Traditional studies on international migration have been conducted either as part of development economics or within the framework of labour economics. The origins of this approach can be traced back to basic neoclassical models establishing a potential for considerable efficiency gains from a more liberal international mobility of labour, (Klein & Ventura, 2007; Moses & Letnes, 2004). Further theories have linked migrants to human capital formation and wages in receiving countries (Massey et al., 1993); but also to financial remittances, education and growth in sending countries (M. Beine et al., 2001; Mountford, 1997; Stark & Wang, 2002).

In parallel, the original neoclassical models also stand as the basic framework to the growing body of theoretical and empirical literature that has explored the role of migrants in favouring transactions between countries. This literature has emphasized the externalities derived from migrant networks in terms of the social and economic linkages between their home and destination countries. These migrant networks externalities act indirectly in reducing informal barriers; and so lowering transaction costs in bilateral economic exchanges between countries. A strand of studies has documented a positive impact of migrant networks on bilateral FDI (Javorcik et al., 2011; Kugler & Rapoport, 2007), firms' internationalization strategies (Foley & Kerr, 2013; Saxenian et al., 2002), international knowledge diffusion (Agrawal et al., 2011; Kerr, 2008) and trade (Gould, 1994).

There has been extensive work on the role played by migrants in boosting bilateral trade (Dunlevy, 2006; Felbermayr et al., 2010; Herander & Saavedra, 2005). Most of these studies use gravity models to assess the pro-trade impact of direct migrant connections between home and host countries along two channels: the preference and the trade-cost ones. The first channel is related to the level of utility migrants might derive from certain goods as compared with others. Thus they will tend to trade more goods from which they get a higher utility in their host countries (Girma & Yu, 2000; Gould, 1994; Head & Ries, 1998; Wagner & Leydesdorff, 2005). The second channel is a self-enforcing mechanism through which migrant networks may help overcoming informal barriers – for instance language, culture and institutions and favour the creation and strengthening of business relationships. Migrants may also carry with them valuable information on foreign business opportunities (Dunlevy, 2006; Herander & Saavedra, 2005). This second channel appears to be the most relevant for explaining other types of international exchanges such as FDI and knowledge, on which migrant connections have been found to have a direct impact.

An interesting development of the literature on migration and trade has explored the role of what we will refer to as indirect migrant connections, which connect minorities from the same origin country across different destinations (Felbermayr et al., 2010; Giovannetti & Lanati, 2015). The seminal work by Rauch & Trindade (2002) is considered as the first empirical work to explore this question, with a special attention on Chinese migrants. The results from this study predict a large indirect trade creation effect of Chinese migrants in their host countries. More precisely, the authors find a large and strong effect: the presence of Chinese population share in two countries, at the levels that prevail in South East Asia leads to an estimated average increase of at least 60 percentage points in bilateral trade in differentiated products between these countries. In general, similarly to the direct migrant connections impacts, the pro-trade effects of indirect migrant networks are not just determined by preferences for certain goods, but also by an alleviation of information frictions. By extension, the literature has investigated other types of international transactions or collaborations enabled by migrant networks, such as knowledge exchanges related to innovation activities.

Linking migration to innovation or international knowledge diffusion has long been considered challenging until the recent development of new global-scale micro data from a variety of sources. This has resulted in an increasing amount of empirical production addressing various issues on the role of migration in innovation or knowledge diffusion in both sending and receiving countries. These studies mainly focus on the specific category of high-skilled migrants, in particular those with degrees or jobs in Science, Technology, Engineering or Mathematics (STEM) (Breschi & Lissoni, 2009; Chellaraj et al., 2008; Hunt & Gauthier-Loiselle, 2010; Kerr, 2009). The most common data sources include labour force surveys and censuses at a national as well as at a global level (Docquier et al., 2007; Docquier & Rapoport, 2012). More recently, a new body of literature has emerged, which uses bibliometric data to track the international mobility of researchers (Appelt et al., 2015; Conchi & Michels, 2014; Kamalski & Plume, 2013; Laudel, 2003; Moed & Halevi, 2014; Moed & Plume, 2013; Pierson & Cotgreave, 2000). Finally, patent data have also been exploited, due to three attractive features:

- first they provide information on homogenous group of high-skilled workers, namely the inventors reported on patent applications;
- second they make it possible to identify migrants by comparing information on the inventors' residence, as reported on the patent documentation, and either their

nationality, which may be also reported on the documentation or inferred with the help of name analysis techniques;

- third, they can help to capture international innovation or knowledge diffusion by means, respectively, of cross-country co-patenting and patent citation analysis.

A study by Miguelez (2016) stands out as a good illustration of this triple advantage of patent data. The author uses inventor data from PCT (Patent Cooperation Treaty) patent records issued by the World International Patent Organization (WIPO). He investigates the impact of hs diasporas on the globalisation of R&D activities. Information on the migrant status of inventors is obtained by comparing the inventors' nationality to their residence at the time of patent filing. As for cross-country collaboration in technology, patent data allow for two measures: co-inventorship and R&D offshoring. The author applies a gravity model with country pairs as observations and either one of the two measures as the dependent variable. The focal regressor is the stock of active hs diaspora from one country i in a host country j , which turns out to have a strong and positive impact on both dependent variables. More precisely, the author finds a 10% increase in the inventor diaspora from i in j leads to an increase of around 2% in international patent collaborations at the level of inventors. This paper represents one major contribution to the empirical literature on the role of direct migrant connections on knowledge flows. However, the focal point in Miguelez' paper remains knowledge diffusion from destination to origin countries and not across destination countries.

In the migration and innovation literature in general, most empirical studies have focused on questions related to the role played by hs diasporas in the diffusion of knowledge to their origin countries (Kerr, 2008) or in enhancing innovation within each specific destination country (Breschi et al., 2014; Chellaraj et al., 2008; Ottaviano & Peri, 2006). This means there are still some unaddressed questions such as how same origin hs diasporas contribute to knowledge diffusion or exchange across destination countries. Besides, this body of literature has been strongly dominated by studies on the US as a host country, although with few exceptions (Niebuhr, 2010; Ozgen et al., 2013).

3.2.2 Science and technology (S&T) collaborations

Bibliometric and patent data have been some of the most widely used data to help investigating collaboration patterns both across individuals – e.g. Scientists, authors and inventors –, institutions, and at a more global scale across countries or regions (Abramo et al., 2012; Kamalski & Plume, 2013; Luukkonen et al., 1993; Narin et al., 1991; Wagner, 2005; Wagner & Leydesdorff, 2005; Yoshikane & Kageura, 2004). Indeed, information recorded in publications on authors and their affiliations have made it possible to develop some measures of scientific collaboration, based upon co-authorship. Besides, publications are one of the most common means of documented scientific communication and collaboration. As such, they have the advantage of covering a wider range of sectors. They therefore appear as an appropriate data source for studies on international collaboration across sectors – university, firm and government – (Glänzel & Schubert, 2004). As for patent data, they are more connected to the industrial R&D. That partly explains why they have been widely used in studies on the link between firms' alliances and innovation, along with data on R&D cooperation (Ponds et al., 2007). This implies each of the above types of data has its own embedded specificities, which may help capturing certain patterns of collaboration better than others. Yet, a common feature to all collaboration patterns is that they are informative of the existence of knowledge networks or linkages that may expand beyond sectors, institutional and national borders, with potential benefits to the involved parties.

Overall, there has been an increasing tendency to international collaborations over the past years worldwide and in OECD countries in particular (Guellec & de la Potterie, 2001). This is due to factors such as massive funding, an increasing mobility of researchers and changes in communication patterns particularly in the scientific world (Glänzel & Schubert, 2004). Also, over the last decades global firms have moved their main technology creation activities from a home-country based perspective to a more internationally-oriented one. This is partly explained by the surge of innovation and rapid trends and requirements of market trends, thus the need to exploit knowledge and technology from different sources abroad (Bastian, 2006; Bresson, 1996; De Backer & Basri, 2008; Nooteboom, 1999; Von Hippel, 1988). Besides these market oriented motivations, there are additional political and institutional factors – such as the European initiative for European collaboration and the Framework Programme for research and technological development at the European level – aiming at promoting

knowledge creation and more particularly R&D cooperation at a regional level (Removille & Clarysse, 1999).

In general firms are the leading players in R&D networks formation or collaboration agreements (De Backer & Basri, 2008). Consequently, studies on the determinants of the globalization of knowledge and technology have long been confined to the firm level (Granstrand et al., 1993). Moreover, incomplete harmonized data on countries' cross-border R&D flows has rendered it difficult to perform cross-countries analysis. However, in recent years new strategies have been used to counter this data shortage. One of these strategies has been to use co-patenting or co-inventorship recorded in patent data at a country level as a proxy for cross-country R&D collaboration. Admittedly, it is important to point out to the singularity of the latter strategy insofar that co-patenting or co-inventorship only representing outcomes from R&D activities, the entire picture might not be fully depicted. Furthermore, unlike other types of S&T collaborations, those captured by patent data mainly involve research collaboration for market purposes and industry-oriented activities (Maggioni et al., 2007). As a result, applied knowledge is often the main outcome of such collaborations (Lata et al., 2012).

Yet, patent data have been increasingly used in various studies on the determinants of international collaboration in innovation. For instance, Guellec & de la Potterie (2001) study the determinants of internationalisation of knowledge at the country level, as proxied by three indicators, namely the shares of patents with the joint presence of: a domestic and a foreign inventor; a foreign inventor and domestic applicant; and a domestic inventor-foreign applicant pairs. They find the internationalisation of a country's technological activity decreases with that country GDP and R&D intensity. This can be interpreted as follows: researchers from large countries prefer to tap into their country's own resources for technology generation. In other words these researchers find it easier to collaborate with other researchers who are closer to them as their high level of technology does not raise any incentive for them to look for any knowledge source elsewhere. The UK and the US are the exception to this rule and Guellec & de la Potterie (2001) explain it with their language similarity with many countries, which thus makes researchers from these countries likely to cooperate more easily with other countries. The latter point implies that language similarity favours the internationalisation of knowledge, as suggested by their results. They also find geographical and technological proximity to foster bilateral cooperation in technology.

As for scientific collaboration, it has been mainly carried out by research institutions such as universities, public and private research centres. The principal motivations of its main actors are among others access to funding and equipment, access to expertise, speeding up progress, enhancing productivity, and reducing isolation (Beaver, 2001). Its study can be traced back to the 1960s (Clarke, 1967). However, most of the research on this topic has used a descriptive approach, at least until recent years (Glänzel et al., 1999; Jones et al., 2008; Katz, 1994; Narin et al., 1991; Okubo & Zitt, 2004; Wagner & Leydesdorff, 2005). Indeed, more recently there have been a growing number of empirical studies on the determinants of international research collaboration using bibliometric data. For instance, Hoekman et al. (2010) who analyze co-publication patterns among European regions for the period 2000-2007. They find patterns of cross-border collaborations to be geographically localized. That is, physical distance impedes on co-publication activities while language similarity favours them.

In general, S&T collaborations – mainly joint publications and joint patents as the ones that are easily traceable with more comprehensive data – have been positively associated with knowledge production and dissemination. For instance, there is a great deal of empirical evidences on the positive impact of R&D alliances on innovation performances (Belderbos et al., 2004; Cincera et al., 2003; Faems et al., 2005). However, most of this literature is at a micro or firm level of analysis and fits more into the scope of management studies, with a special attention on company or university strategies.

All in all, the above theoretical and empirical literature review reveals although his migration has been the topic of a large empirical and theoretical amount of work, linking it to international collaboration in knowledge is a new approach with several questions which are yet to be explored. We therefore intend to contribute to this field of literature by digging into some of these questions.

3.3 Key definitions

Before going further into analysis, we need to introduce the key definitions we will use in the present study. We make a distinction between general definitions – those derived from the migration literature – and specific definitions – which we have elaborated ourselves for our own research purposes.

- **General definitions**

In general, by “migrants” we refer to individuals residing outside their countries of origin at a given point of time. In turn, the migration literature alternatively defines the country of origin according to three non-mutually exclusive criteria: nationality, place of birth and duration of stay. There are specific limitations and advantage tied to each criterion, and often time the choice between them depends on data availability. Our data – a detailed description of which is provided below – come from national labour force surveys and censuses, which mainly refer to migrants’ place of birth, with the exception of a couple of sources which use the nationality instead. In general, they do not distinguish between temporary and permanent migration in our analysis²⁷.

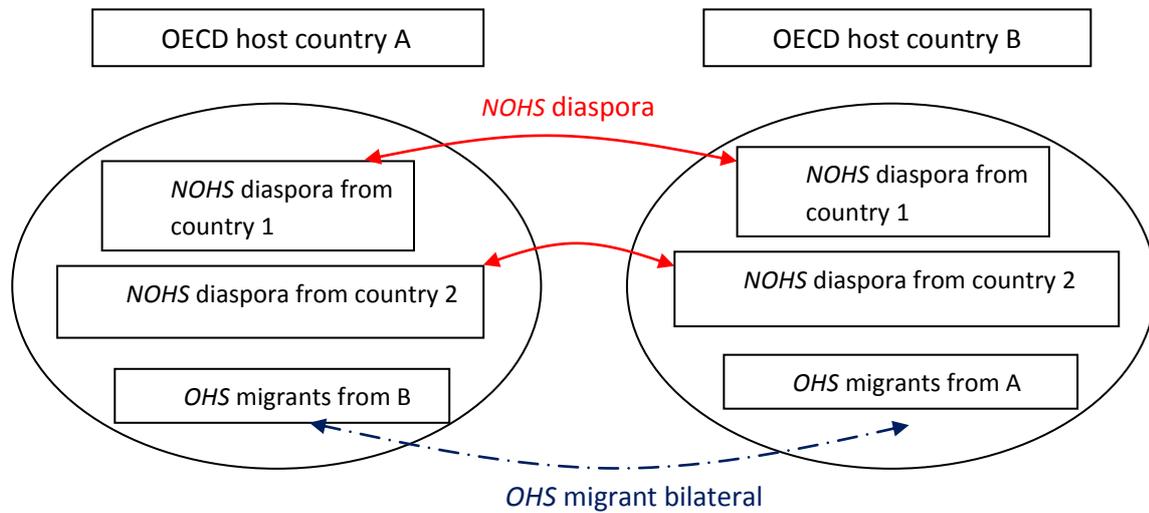
Secondly, drawing from the conceptual definition of the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the OECD Frascati Manual, we identify highly-skilled migrants on the basis of their educational attainment level. More precisely, we define as hs individuals having completed a tertiary education level.

- **Specific definitions**

Given the research objectives of this paper, we adopt the following key definitions, as illustrated by Figure 3.1. By non-OECD hs (NOHS) diaspora – migrants from countries 1 and 2 in our illustration –, we mean all hs migrants whose country of origin does not belong to the OECD host countries – represented by country A and B in Figure 3.1. We are particularly interested about NOHS groups of people from the same country of origin, but dispersed across two or more OECD host countries. Migrants in such groups might interact with each other across borders and form connections – as represented by the red bold arrows in Figure 3.1. These connections differ from the OECD hs (OHS) migrant bilateral links formed by groups of hs migrants originating from and moving to OECD host countries, as depicted by the blue broken arrows.

²⁷Notice that the United Nations (UN) has recommended considering as migrants only those with at least a one-year stay abroad. We acknowledge this recommendation, but our data does not allow us to take into account any duration of stay. However, we do believe not accounting for the time dimension of mobility wouldn’t bias our results, as we assume the duration of stay has a minor role to play in the formation of networks

Figure 3.1: NOHS diasporas Vs. OHS migrant bilateral links.



Besides, we define S&T collaborations the joint efforts of groups of organisations in performing R&D activities (Lata et al., 2012). Here we explore two types of S&T collaborations at an international dimension; co-inventorship and co-authorship.

3.4 Methodology

3.4.1 Empirical approach

For assessing the contribution of NOHS diasporas to international collaboration in S&T, we use a gravity model, which explains the intensity of interaction between two elements with both their individual characteristics and mutual distance. In social sciences, and economics in particular, it has been widely applied at the regional or country level to explain both trade volumes (Bergstrand, 1985; Felbermayr & Toubal, 2012; Rauch, 1999; Tinbergen, 1962), migration (Beine, et al., 2016; Mayda, 2010; Ortega & Peri, 2013) and knowledge diffusion (MacGarvie, 2006; Maggioni et al., 2007; Miguelez, 2016; Miguelez & Noumedem, 2017).

Our analysis takes place at the country pair level²⁸, as follows:

$$KC_{ij} = e^{\alpha_0} \cdot NOHSShare_{ijk}^{\alpha_1} \cdot Z_{ij}^{\gamma_n} \cdot e^{\tau_i} \cdot e^{\tau_j} \cdot \varepsilon_{ij} \quad (3.1)$$

Where

- KC_{ij} stands alternatively for one of our two dependent variables, namely co-inventorship and co-authorship between OECD countries i and j ;
- $NOHSShare_{ijk}$ is the product of the share in percentage of hs migrants from non-OECD country k residing in countries i and j respectively; (notice that all our country pairs consist of destination countries, with migrants coming from outside these destination countries)
- Z_{ij} is a set of n dyadic covariates and country specific control variables;
- τ_i , and τ_j are country i and country j fixed effect (FE) respectively;
- ε_{ij} is the error term.

Both in the trade or migration literature, due to the presence of many zeros in the dependent variable the multiplicative form of the gravity equation is often transformed into its logarithmic form before being regressed with an Ordinary Least Square (OLS). However, this may induce heteroskedasticity in the error terms, hence some inconsistency of estimation. As a remedy, we follow Santos-Silva & Tenreyro (2006) methodology and estimate the gravity equation using a Poisson pseudo-maximum likelihood (PPML) model. One advantage of PPML estimations is that when the dependent variable is a count variable and the covariates are in logarithms, coefficient estimates can be interpreted as elasticities. We therefore transform equation (1) into its conditional expectation form as follows:

$$E(KC_{ij} | X_{ij}) = \exp[\alpha_0 + \alpha_1 \ln NOHSShare_{ijk} + \gamma_n \ln Z_{ij} + \tau_i + \tau_j + \varepsilon_{ij}] \quad (3.2)$$

We add a unit to all explanatory variables in order to correct for the zero values in the natural logarithmic transformation. Following Felbermayr et al. (2010), we run cross-section regressions for each period.

²⁸ We run cross-section regressions at two separate points of time corresponding to five-year windows: 2000 – 2004 and 2010 – 2014 respectively. Hence we do a five-year average of all variables except for time invariant variables.

In further model specifications, we introduce the OHS migrant bilateral links variable – $OHSBShare_{ij}$ – in the baseline model in order to control for the migration effects from OHS migrants themselves. This variable is the product of the percentage of hs migrants from country i living in country j and the percentage of hs migrants from country j living in country

3.4.2 Data

We consider as destinations 31 OECD member countries (see Table B1 in Appendix B) out of 35, as of 2017²⁹. They account for the largest share of cross-border flows of R&D worldwide (De Backer & Basri, 2008) and include the top five hs migrants host countries in the world. As for the origin countries, we focus on India and China, which are the top non-OECD countries of origin for hs migrants (Widmaier & Dumont, 2011) (see also Table B6 in Appendix B)³⁰. Additionally, India and China are among the fastest growing migrant countries of origin.

Our observations are country pairs. For the dependent variable – cross-country scientific and technology collaboration – we consider two alternative proxies: co-inventorship and co-authorship. Each of them captures different features of the phenomenon of interest. We examine them in turn.

3.4.2.1 Dependent variables

We define co-inventorship by considering inventors collaboration on a single patent. This helps capturing the overall joint inventorship between countries, regardless of the different standing of partners. This implies that co-inventors may or may not come from the same company – within Multinational Corporations (MNCs). Therefore although computed at the individual level, the co-inventorship variable may point to links between individuals, organizations or individual-organization. The latter is one important feature which differentiates the co-inventorship variable from the co-authorship one as it will be seen below.

We compute the co-inventorship variable from raw data. In particular, we use the February 2015 version of the OECD REGPAT database, which covers patent applications to the

²⁹ We exclude Turkey due to some data inconsistency in foreign migration records. We also exclude South Korea, Latvia and Iceland, for which we have no or severely incomplete migration figures.

³⁰ Indeed, the other top sending countries are largely EU countries; which generates a mix of intra-regional – intra-EU – and international – EU-US or EU-Canada – migration patterns.

European Patent Office (EPO) – as derived from PatStat, the Worldwide Patent Statistical Database released in Autumn 2014 – and PCT patents – the patents database from the World International Patents Office (WIPO) (OECD, REGPAT database, Feb., 2015)³¹. This database records information on patent inventors and their addresses. We check for duplicates on patents application in both PCT and EPO data on the basis of their application number, applicant names and addresses and remove them. For the international co-inventorship variable we only keep patent records with at least two co-inventors with their respective addresses in two distinct OECD countries. Then we proceed to counting co-inventorships by pair of countries, by year. Each patent generates as many co-inventorship counts as the number of combinations of OECD countries we can obtain from the inventors’ addresses. The more inventors from two OECD countries are listed on the same patent, the higher the count. We do not record the combinations involving one or two non-OECD countries. Tables 1a and 1b provide an example: patent WO2005100777 reports three inventors from Germany and as many from Austria, which generates 9 Germany-Austria co-inventorships ; patent EP1320536 reports three inventors from Switzerland and one from France, which generates 6 Switzerland-France co-inventorships ; and so forth.

Table 3.1a: Example of co-inventorship data (two patents)

Patent id	Inventors’ Country	Year
WO2005100777	AT	2004
WO2005100777	AT	2004
WO2005100777	AT	2004
WO2005100777	DE	2004
WO2005100777	DE	2004
WO2005100777	DE	2004
WO2005100777	IT	2004
WO2005100777	IT	2004
EP1320536	CH	2000
EP1320536	CH	2000
EP1320536	CH	2000
EP1320536	FR	2000
EP1320536	GB	2000

Table 3.1b: Count of co-inventorships, by country pairs from data in Table 3.1a

Ctry1	Ctry2	Co-inventorships	Year
AT	DE	9	2004
AT	IT	6	2004
DE	IT	6	2004
CH	FR	3	2000
CH	GB	3	2000
FR	GB	1	2000

³¹See Squicciarini et al. (2013) for a detailed description of the database.

Our co-authorship variable measures the extent of collaboration for scientific or basic knowledge production purposes. Publication data cover a wider range of fields than patent data. Hence, the co-authorship variable stands as a broader scope proxy for collaborations. Additionally the co-authorship variable fully captures inter-institutions collaboration, insofar that international collaborations recorded in publications are commonly made between authors affiliated to different institutions; which is not the case in patents where intra-company collaborations are more frequent. The co-authorship data were provided by the OECD, based on elaborations of Scopus Custom Data³². The co-authorship variable is built at the country-pair level, based on information on authors' country of affiliation as recorded on publication. An affiliation can be a university or any research institution. Thus the co-authorship variable is just the sum of co-authorships per country pair, per year. Similarly to co-inventorship, we count each co-authored publication across each country pair as single instance – one unit – of co-authorship, no matter the number of authors involved. We only consider co-authorships among the OECD countries of our sample.

3.4.2.2 Explanatory variables and controls

In order to track the NOHS diasporas, we need information on hs migrants from China and India to OECD destination countries. Additionally, we need data on the bilateral hs migration between OECD countries so to compute the OHS migrant bilateral links variable.

Hence, we use the first and third editions of the OECD-DIOC database (Database on Immigrants in OECD Countries; 2000/01, 2010/11)³³, which assembles information from various national sources on the stock of immigrants for each of the OECD destination countries, from around 200 origin countries. Table B1 in Appendix B provides details on the data source for each destination country.

³²That is a customised large-scale dataset for research performance analysis derived from Scopus core records. Scopus itself is a bibliometric database owned by Elsevier which contains abstracts and references from over 22,748 peer-reviewed journals from 4,000 publishers worldwide; thus a wider multidisciplinary coverage.

³³There are three editions of DIOC: 2000/01, 2005/06 and 2010/11, which are demographic data collected on 34 OECD countries based on national censuses, population registers and labour force surveys conducted over a given period approximately throughout the sample of destination countries – between 1999 and 2003 for DIOC 00/01 and 2010 to 2013 for DIOC 10/11. In parallel, there is an extension of DIOC 00/01, DIOC-E which includes 66 additional non-OECD destination countries. We only work with DIOC 00/01 and 10/11 as in the DIOC 05/06 data for many receiving countries like UK, information on immigrants' origin has been aggregated at the regional level, which makes it impossible to identify immigrants from Indian or Chinese origin. It is therefore not possible to get figures on Indian and Chinese hs migrants in these OECD destination countries for the period covered by DIOC 05/06.

Immigrants are mostly identified on the basis of their place of birth³⁴. Additionally, DIOC reports information on the migrants skill level, as proxied by their educational attainments. We consider as highly-skilled all the individuals who have completed a tertiary level of education; in other words, those belonging to levels 5 and 6 according to the International Standard Classification of Education (ISCED) of UNESCO. We compute the NOHS diaspora variable ($NOHSShare_{ijk}$) as the product of the percentage of hs migrants from India/China in the total hs population, for each pair of OECD receiving countries. As for the OHS migrant bilateral links variable ($OHSBShare_{ij}$) for each OECD country pair we compute the product of the percentage of hs migrants from one country and residing in the other country of the pair.

The NOHS diaspora variable, our focal explanatory variable, is computed as below:

$$NOHSShare_{ijk} = \frac{H_{ik} * H_{jk}}{H_i * H_j} \quad (3.3)$$

where H_{ik} and H_{jk} are the hs migrants population from country k – with k equivalent either to China or India – respectively in OECD countries i and j, while H_i and H_j are the total hs population in these same countries.

The OHS migrant bilateral links variable is computed as followed:

$$OHSBShare_{ij} = \frac{H_{ij} * H_{ji}}{H_i * H_j} \quad (3.4)$$

where H_{ij} and H_{ji} are respectively the total hs migrant population from country i in country j, and vice versa.

Our main controls consist of a set of dyadic and country-specific variables that account for the effect of other factors affecting the intensity S&T collaboration.

At a dyadic level, we control for the physical and cultural distance between country pairs with several variables from the ‘Centre d’Etudes Prospectives et d’Informations Internationales’ (CEPII)³⁵. First, we consider two variables for physical distance. One is a dummy variable for contiguity, taking value 1 if the two countries in the observation share a common border and 0 otherwise. The other one is a variable measuring the distance in kilometers between the

³⁴Except for Germany and Japan where only information on nationality is available.

³⁵A detailed description of these variables can be found in Mayer & Zignago (2011).

biggest cities of both countries, weighted by the share of these cities' population in the overall country's population. Second, cultural ties are proxied with a dummy for common language taking value 1 if a language is spoken by at least 9% of the population in both countries and 0 otherwise. Last, for historical ties we include a dummy taking value 1 if there has been a past colonial link between both countries and 0 otherwise. Additionally in the co-inventorship equation, we include an index of technological proximity that controls for the similarity of technological specializations of the two countries. This index is computed using patents data from the CRIOS–Patstat database, which is derived from the 2014 EPO patent database (Coffano & Tarasconi, 2014). In particular:

$$Tech. similarity_{ij} = \frac{\sum f_{ih} f_{jh}}{(\sum f_{ih}^2 \sum f_{jh}^2)^{1/2}} \quad (3.5)$$

where f_{ih} stands for the share of patents of technological class h – according to 30-class reclassification of IPC codes³⁶ - of country i , and f_{jh} the share of patents of technological class h of country j . Values of the index close to the unity indicate that countries of a given pair are technologically similar, and values close to zero mean they are technologically far from each other (Jaffe, 1986).

At the country level, we add technological masses – which in the gravity model help testing for attraction level between both countries – that control for countries specific characteristics in terms of science and technology intensity or capability. For the co-inventorship equation, we interact the five-year averages for each of the three waves – 2000-04, 2005-09, 2010-14 – of the total patents in countries i and j . That is, we calculate the product of the five-year average of total patents in the two countries. The latter figure comes from the 2015 WIPO statistics database. It is the total number of patents per applicants origin filed at the European Patent Office (EPO)³⁷. Similarly, we add an interaction term for the five-years averages of publications in countries i and j in the co-authorship equation. These total publications are

³⁶This 30-class reclassification of IPC codes was originally proposed by the OST (Observatoire des Sciences et Techniques). For more details see Coffano & Tarasconi (2014).

³⁷Using this data, one might think of a potential bias occurring for an over-representation of European patents in the EPO – like Germany for instance – as compared with patents from other parts of the world. However, we doubt this actually biases our results for several reasons. First most OECD countries are European, and for the rest of countries, except for the US, Japan and Israel their inventive activities is comparatively insignificant. Secondly, assuming we have overestimated the technological mass of European countries, the results from table 3.6 show the potential over estimation might not be high. If we look at the patents product variable estimates, in all model specifications without the US (as one of the assumed underestimated weight), this variable coefficient drops drastically (sometimes by over half). This implies the weight of the US is still very much important. As for Japan, there aren't many co-inventorship instances with that country.

derived from the SCImago Journal & Country Rank³⁸ – see SCImago, 2007 –, which we obtained from the Scopus database.

Last but not the least, we control for economic masses by adding the product of the five-year averages of GDPs for the countries in each pair of countries *i* and *j*.

Table 3.2 Variables definition

Variables	Definition
Dependent variable	
Co-inventorship	Count of the number of joint patents inventorship instances between countries <i>i</i> and <i>j</i> .
Co-authorship	Count of the number of joint publications authorship instances between countries <i>i</i> and <i>j</i> .
Explanatory and control variables	
NOHS diaspora	Product of the percentage of NOHS migrants from same origin in countries of destination <i>i</i> and <i>j</i> . In other words, it is the probability that if we select an individual at random from each of the two countries' total hs population, both will be from the same NOHS migrant country of origin.
OHS migrant bilateral links	Product of the percentage of OHS migrants from country <i>i</i> living in country <i>j</i> and the percentage of OHS migrants from country <i>j</i> living in country <i>i</i> . Or it is the probability that, if we select an individual at random from each of the two destination countries total hs population, both will be hs migrants from one of the two countries living in the other.
Technological similarity	Index of technological proximity between countries <i>i</i> and <i>j</i> varying from 0 to 1.
Contiguity	Dummy variable for taking value 1 if countries <i>i</i> and <i>j</i> share same border and 0 otherwise.
Colony	Dummy variable taking value 1 if both countries <i>i</i> and <i>j</i> share a common past colonial history and 0 otherwise.
Common official language	Dummy variable taking value 1 if same language is spoken by at least 9% of the population in countries <i>i</i> and <i>j</i>
Distance	Distance in kilometres between the biggest cities of both countries <i>i</i> and <i>j</i> , weighted by the share of these cities' population in the overall countries' population
Product of patents in countries <i>i</i> and <i>j</i>	Product of the total number of patents in countries <i>i</i> and <i>j</i> ; #Pat _{<i>i</i>} *#Pat _{<i>j</i>}

³⁸SCImago Journal & Country Rank is an online portal that reports journals and country scientific indicators developed from information recorded in SCOPUS. Both data sources are comprehensive at the geographical and thematic levels in a sense that they cover a large range of countries worldwide and a wide interdisciplinary content.

Product of publications in countries i and j Product of the total number of publications in countries i and j; #Pub_i*#Pub_j

Product of GDP in countries i and j Product of the GDP in countries i and j; #GDP_i*#GDP_j

3.4.3 Descriptive statistics

Based on the 31 OECD receiving countries in our sample, we have 465 observations per DIOC edition (DIOC 00/01 and 10/11) and per immigrants' country of origin (China and India).

Table 3.3 below depicts the top fifteen S&T collaboration corridors between 2010 and 2014. More precisely, it shows the five-years averages of co-inventorship and co-authorship for the fifteen country pairs with the highest figures.

One interesting evidence from the table is the leading position of the US as a key research partner for many European and non-European countries. Indeed, the country is present in ten country-pair collaborations, in both patenting and publication out of the top fifteen collaborations of our table.

Table 3.3: Science and technology research corridors for years 2010 – 2014

Country a	Country b	Average co-inventorship*	Country a	Country b	Average co-authorship*
USA	Germany	3,388	USA	UK	22,334
USA	UK	2,629	USA	Germany	19,865
USA	Canada	2,544	USA	Canada	18,687
Germany	Switzerland	1,572	USA	France	13,013
France	Germany	1,397	UK	Germany	11,693
USA	France	1,201	USA	Italy	11,366
USA	Japan	1,171	USA	Japan	10,116
USA	Switzerland	927	USA	Australia	9,898
USA	Israel	804	France	Germany	8,615
Germany	Austria	717	UK	France	8,430
UK	Germany	679	USA	Spain	8,353
France	Switzerland	641	USA	Netherlands	7,930
USA	Netherlands	604	Italy	UK	7,514
USA	Belgium	571	Germany	Switzerland	7,221
USA	Italy	568	USA	Switzerland	7,180

* This number corresponds to the five-year average of S&T collaborations between 2010 and 2014

The Chinese hs migrant figures from DIOC 10/11 and DIOC 00/01 are reported in Table 3.4 below. The US stand out as the country with the biggest intake of Chinese hs in absolute terms in both DIOC editions; from a total number of over 372,000 Chinese hs in the first wave to over 484,000 Chinese hs in the third edition, so an increase of nearly 30%. However, this increase is smaller than that of all other receiving countries, where the total number of Chinese hs has more than doubled or tripled over the same period of time. For instance in Canada – which is the second top receiving country of the list – the intake of Chinese hs has gone from over 86,000 to 208,000, an increase of almost 142%. A glance at the columns reporting the shares of Chinese hs on the total hs population in each receiving country shows a similar story. The share of Chinese hs migrants over total hs from DIOC 00/01 to DIOC 10/11 has increased in all receiving countries. Interestingly, Australia appears as the country with the biggest shares of Chinese hs in the total hs population with 1.43% and 3.13% in the first and third editions respectively.

Table 3.4: Top fifteen receiving countries of Chinese hs migrants in DIOC 10/11 and 00/01

DIOC 10/11				DIOC 00/01			
Country of destination	Chinese hs immigrants ('000)	Total hs residents ('000)	Chinese share of hs (%)	Country of destination	Chinese hs immigrants ('000)	Total hs residents ('000)	Chinese share of hs (%)
USA	484.2	59,088	0.82	USA	372.8	46,304	0.81
Canada	208	9,155	2.27	Canada	86	6,320	1.36
Australia	116.2	3,713	3.13	Japan	43.9	21,125	0.21
UK	76.5	11,827	0.65	Australia	32	2,252	1.43
Japan*	75.1	-----	----	UK	11.7	6,856	0.17
New Zealand	25.1	1,018	2.47	Germany	6.2	8,540	0.07
Germany	22.9	11,610	0.20	France	5.3	6,305	0.08
France	16.9	9,717	0.17	New Zealand	5	567	0.88
Spain	7	8,514	0.08	Sweden	1.8	1,137	0.16
Italy	4.6	4,797	0.10	Switzerland	1.7	824	0.20
Sweden	4.6	1,575	0.29	Spain	1.4	4,266	0.03
Switzerland	4.2	1,439	0.29	Italy	1.3	3,004	0.05
Netherlands	3.6	2,763	0.13	Belgium	1	1,432	0.07
Ireland	2.8	786	0.35	Austria	0.7	582	0.12
Austria	2.0	809	0.25	Ireland	0.6	562	0.11

*For Japan we were not able to compute total hs figures from the DIOC 10/11 data as all education levels for Japanese nationals were recorded as unknown.

The figures for the top Indian hs receiving countries are very similar to those for the Chinese, except for a few cases. In Table 3.5, the US appear to be the top receiving country of Indian hs in absolute terms in both DIOC editions – from over 504,000 to nearly 975,400. In DIOC 10/11 figures, Canada loses to the UK its second position occupied in DIOC 00/01 – from a total Indian hs of over 101,100 to 230,400 for Canada and from 96,900 to 269,400 for the UK. As for the shares of Indian hs in the total hs population, they have increased in all countries from the first to the third DIOC edition. Canada and Australia are the countries with the highest shares in the first and third editions respectively – 1.60% and 4% respectively.

Table 3.5: Top fifteen receiving countries of Indian hs migrants in DIOC 10/11 and 00/01

DIOC 10/11				DIOC00/01			
Country of destination	Indian hs immigrants ('000)	Total hs residents ('000)	Indian share hs (in%)	Country of destination	Indian hs Immigrants ('000)	Total hs residents ('000)	Indian share hs (in%)
USA	975.4	59,088	1.65	USA	504	46,304	1.09
UK	269.4	11,827	2.28	Canada	101.1	6,320	1.60
Canada	230.4	9,155	2.52	UK	96.9	6,856	1.41
Australia	148.5	3,713	4.00	Australia	34	2,252	1.51
New Zealand	30.7	1,018	3.01	New Zealand	5.8	567	1.02
Germany	9.9	11,610	0.09	France	3.9	6,305	0.06
Ireland	9.8	786	1.25	Israel	2.4	1,036	0.23
France	8.4	9,717	0.09	Switzerland	2.4	824	0.29
Switzerland	7.2	1,439	0.50	Japan	1.9	21	0.01
Italy	5.6	4,797	0.12	Netherlands	1.9	1,911	0.10
Sweden	4	1,575	0.25	Sweden	1.7	1,137	0.14
Israel	3.7	1,531	0.24	Italy	1.6	3,004	0.05
Netherlands	3.1	2,763	0.11	Ireland	1.5	562	0.26
Spain	3	8,514	0.04	Belgium	1.5	1,432	0.10
Japan*	2,9	-----	----	Norway	1	640	0.17

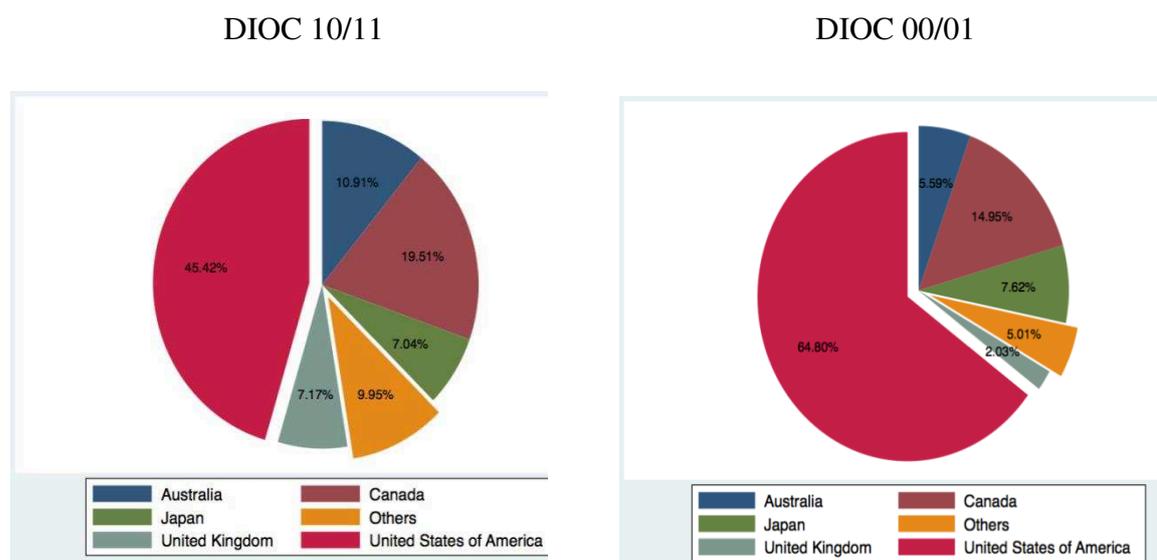
*For Japan we were not able to compute total hs figures from the DIOC 10/11 data as all education levels for Japanese nationals were recorded as unknown.

Overall, these figures show a skewed distribution of Indian and Chinese hs migrants across countries, as these migrants concentrated mostly in the top four countries of the list. These distributions are better illustrated in the following graphs.

In Figure 3.2, we see the change in the distribution of Chinese hs migrants in DIOC 10/11 and DIOC 00/01 respectively for the OECD countries of our sample. These graphs show the

biggest intakes of Chinese hs to be in five countries only, out of the 31 countries in our sample. These countries are the US, Canada, Japan, Australia and the United Kingdom, whose total share of Chinese hs accounts for around 90.05% and 94.99% of the total Chinese hs migrants in our sample, in DIOC 10/11 and DIOC 00/01 respectively. However it is important to mention the differences between the two graphs, with the US losing its shares of Chinese hs – from 64.80%, to 45.42% - to the benefit of other destination countries. In other words, Chinese hs migrants seem to find other destination countries such as Australia and the UK increasingly more attractive. This might be due to factors such as the loosening of emigration controls in China (Ortega & Peri, 2009), but also the gradual shift of immigration policies in some European countries like in France, away from their traditional focus on family reunions and asylum seeking to more hs-oriented policies (Docquier et al., 2007).

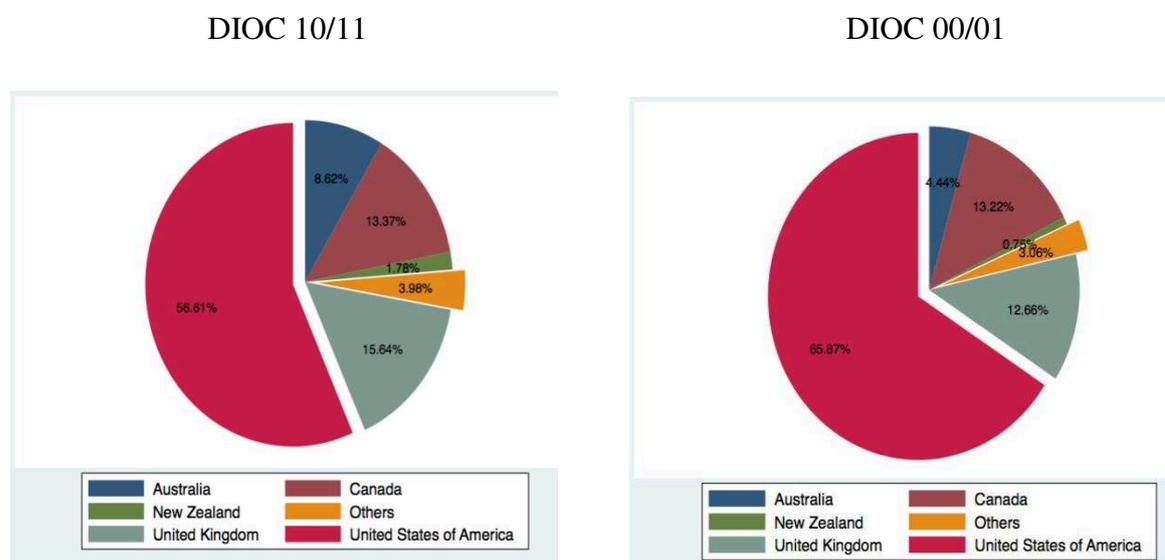
Figure 3.2: Chinese hs migrants distribution in 31-OECD destination countries: DIOC 10/11 vs. DIOC 00/01



Similar remarks can be made from the figures depicting the distribution of Indian hs migrants in our sample of receiving countries as shown in Figure 3.3 below. The distribution tends to be slowly becoming more even from DIOC 00/01 on the right hand side to DIOC 10/11 on the left. But this process is seemingly confined to the group of the four biggest receiving countries; the US, Canada, the United Kingdom and Australia. Hence, this implies some clustering effect of the Indian hs diaspora in these countries. Interestingly, an important share of these Indian hs diaspora is made of skilled professionals, business scientists and academic

elites (Docquier & Rapoport, 2012), and these countries seem to be more attractive to these countries.

Figure 3.3: Indian hs migrants distribution in DIOC 10/11 and DIOC 00/01



3.5 Results

3.5.1 The effect of the Chinese and Indian hs diaspora in S&T collaborations

Table 3.6 below reports the results from the co-inventorship regressions for Chinese and Indian hs diasporas for DIOC 10/11 and DIOC 00/01 respectively, with country and time fixed effect. The coefficient estimates are elasticities. The results for the Chinese hs diaspora are illustrated in columns (1) to (6) – with columns (1) to (3) showing DIOC 00/01 results and columns (4) to (6) showing those from DIOC 10/11 -, while columns (7) to (12) present the outcomes from the Indian hs diaspora in DIOC 00/01 - (7) to (9) – and DIOC 10/11 respectively - (10) to (12). For each NOHS diaspora variable and for each DIOC edition, we use three model specifications.

In the baseline model we only include our main variable of interest, $\ln\text{NOHSSshare}_{ijk}$, along with basic control variables. We then add the OHS migrant bilateral links variable $\ln\text{OHSBShare}_{ij}$ to control for the effect of hs migrants exchanges between host countries i and j . Last, as a robustness check, we drop from our sample all country pairs that include the US–

so to check whether it is that specific country which drives our results, due to its double status as both the top hs migrants host country and the most recurrent partner in international R&D collaborations, within and across firms.

From the baseline regressions for the Chinese hs variable – columns (1) and (4) –, we get different results for the elasticities of our main variable of interest in DIOC 00/01 and DIOC 10/11 respectively. Indeed, while for the first DIOC edition we get a strong positive and statistically significant result – 1.321 –, for the second edition the coefficient value decreases to 0.285. In other words we find that, if we double the probability of getting Chinese migrants from a random draw of two individuals from two host countries total hs population, co-inventorship between these two countries will increase by 132.1% in DIOC 00/01 and by 28.5% in DIOC 10/11. Projecting these results into our sample average figures (Table B2 in Appendix B) gives us the marginal effects. Our sample average probability of having two Chinese hs in a country-pair random draw is 0.041 in DIOC 00/01. If that number doubles to 0.082 the induced effect will be an important raise in the average number of co-inventorships among country pairs to the value of 223³⁹. By analogy, if we double the DIOC 10/11 sample average probability of Chinese hs from 0.987 to 1.974, this will induce a change in the average number of country pair co-inventorships from 72 to almost 93⁴⁰. Controlling for the OHS migrant bilateral links variable $\ln\text{OHSBShare}_{ij}$ in the baseline equations only modestly changes the results. As shown in columns (2) and (5) for the first and third DIOC editions respectively, our main coefficient remains positive and statistically significant but slightly drops to 1.284 and 0.269 in DIOC 00/01 and 10/11 respectively. In parallel we get positive and significant results for the OHS migrant bilateral links variable - 0.991 and 0.758 respectively. Interestingly, the coefficient for the Chinese hs variable is stronger than the one for the OHS migrant bilateral links in the DIOC 00/01. When dropping the US as a destination country from our regressions, the magnitude of our main variable coefficient decreases to 0.950 and 0.218 in first and second DIOC editions as shown in columns (3) and (6)

³⁹Tables A2 in the Appendix reports the average number of country-pair co-inventorships in DIOC 00/01 for our sample to be 96. Applying a 132.1% increase to that number we get $127 + 96$ which gives 223. In general, we use this rule to get the marginal effect of each of our explanatory variables based on our sample mean values for each of the dependent variables. That is, we get the marginal effect in terms of each of our dependent variables as $N(\alpha + 1)$, where N is the average value of that variable in our sample and α is the coefficient estimate of the migration variable

⁴⁰ It is important to note that this number represents the change in the sample average co-inventorship in absolute terms and not the change in the average number of co-patenting. Indeed, our co-inventorship was built by counting all occurrences of country pair inventors' residency in a single patent. Therefore, this change might as well capture an increase of co-inventorship within patents; or an increase of patents co-invented across a country-pair.

respectively. As for the coefficients of the usual gravity covariates, they are in line with findings from the trade or migration literature except for a few cases where we get insignificant results.

Two expected results to stress are the positive and significant coefficients of the variable for the product of total patents in each of the two countries and the variable for technological similarity in all model specifications. The positive and significant sign of the total patents product variable suggests that countries that file more patents are more likely to get involved in co-inventorship activities. While the result of technological similarity suggests that countries performing in technologically similar fields have a greater chance to engage in joint inventorship.

The outcomes for India as an origin country are similar to those for China. The baseline model of DIOC 00/01 and 10/11 in columns (7) and (10) respectively return positive and significant results for the coefficients of our main variable of interest – 1.008 and 0.298. This means, if the probability of getting two Indian hs from a random draw of two individuals from the total hs population of the two host countries doubles, the co-inventorship between these countries would increase by 100.8% in DIOC 00/01, and by 29.8% in DIOC 10/11. Following a similar reasoning as earlier, we translate these results into their marginal effect values by projecting them into our sample average figures from Table B2 in Appendix B. Doubling the DIOC 00/01 sample average probability of having two Indian hs in a country-pair random draw from 0.071 to 0.142 would induce an increase in the average co-inventorships from 96 to nearly 193. Similarly, in DIOC 10/11 an increase in the average probability of Indian hs from 0.331 to 0.662 leads to an increase in the co-inventorships sample mean from 72 to 94. When adding the OHS migrant bilateral links variable to the baseline equation, our main coefficient remains positive and significant as illustrated in columns (8) and (11) respectively – 1.028 and 0.605. Removing the US as a destination country from the data doesn't affect our results sign. We still get positive and significant coefficients, although their value diminishes – 0.448 and 0.249 as shown in columns (9) and (12) respectively. This suggests Indian hs diaspora seem to have a major impact on co-inventorship activities between the US and the rest of OECD countries. Other control variable results do not differ much from what we had for the case of China as described above. Again here we get positive and significant results for the estimates of the product of total patents and technological similarity index.

Table 3.6: Chinese and Indian hs diaspora in international co-inventorship

Co-inventorship	DIOC 00/01 CHINA			DIOC 10/11 CHINA			DIOC 00/01 INDIA			DIOC 10/11 INDIA		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
lnNOHSShare _{ijk}	1.321** (0.638)	1.284** (0.637)	0.950** (0.401)	0.285** (0.126)	0.269*** (0.093)	0.218** (0.102)	1.008*** (0.273)	1.028*** (0.283)	0.448* (0.260)	0.298* (0.163)	0.605*** (0.186)	0.249** (0.118)
lnOHSBShare _{ij}		0.991*** (0.218)	0.694*** (0.179)		0.758*** (0.148)	0.650*** (0.219)		0.937*** (0.286)	0.461*** (0.148)		0.745*** (0.286)	0.550** (0.227)
Common lang.	0.558*** (0.196)	0.344** (0.170)	0.467*** (0.163)	0.469*** (0.099)	0.377*** (0.083)	0.434*** (0.136)	0.512** (0.235)	0.346 (0.211)	0.476** (0.192)	0.609*** (0.191)	0.499*** (0.178)	0.361*** (0.1228)
ln(distance)	-0.346*** (0.109)	-0.403*** (0.104)	-0.371*** (0.070)	-0.117 (0.094)	-0.095 (0.071)	-0.093 (0.080)	-0.322*** (0.111)	-0.348*** (0.106)	-0.287*** (0.084)	-0.329** (0.097)	-0.239** (0.094)	-0.337*** (0.057)
Contiguity	0.221 (0.171)	-0.082 (0.196)	0.315** (0.128)	0.593*** (0.128)	0.398*** (0.122)	0.465*** (0.125)	0.274 (0.181)	-0.024 (0.215)	0.310** (0.139)	0.303** (0.135)	0.108 (0.169)	0.350*** (0.122)
Colony	0.431** (0.201)	0.325* (0.179)	-0.166 (0.202)	-0.072 (0.100)	-0.148 (0.090)	-0.225 (0.149)	0.170 (0.206)	0.045 (0.220)	-0.022 (0.235)	0.205 (0.168)	0.085 (0.190)	-0.217 (0.138)
Tech. similarity	3.400* (1.806)	2.902* (1.715)	2.057* (1.158)	3.193*** (0.717)	3.254*** (0.620)	2.736*** (0.601)	3.413* (2.022)	3.176* (1.880)	1.443 (1.248)	3.103** (1.490)	3.011** (1.472)	3.074*** (0.643)
ln(GDP _i *GDP _j)	1.296** (0.590)	1.524*** (1.505)	-0.412 (0.309)	0.615 (1.267)	-0.619 (0.489)	0.830*** (0.310)	1.745*** (0.654)	1.834*** (0.616)	0.243 (0.219)	2.033*** (0.543)	2.340*** (0.615)	0.830*** (0.230)
ln(patent _i *patent _j)	1.094*** (0.152)	1.160*** (0.143)	0.640*** (0.093)	1.110*** (0.175)	0.985*** (0.063)	0.203*** (0.059)	0.759*** (0.088)	0.796*** (0.080)	0.520*** (0.069)	0.850*** (0.061)	0.803*** (0.047)	0.458*** (0.041)
Constant	-39.26*** (13.28)	-44.25*** (11.25)	3.669 (6.119)	-30.60 (29.87)	-2.602 (10.68)	-19.96*** (6.649)	-43.30*** (13.15)	-45.57*** (12.56)	-8.124*** (4.004)	-53.19*** (11.93)	-60.78*** (13.51)	-21.33*** (5.072)
Observations	465	465	435	465	465	435	465	465	435	465	465	435
Pseudo R2	0.943	0.964	0.959	0.987	0.993	0.976	0.925	0.945	0.965	0.931	0.947	0.976
Countries & time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Without the US	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

We add a unit to all of the above explanatory variables before logarithmic transformation in order to account of the presence of many zeros

For the co-authorship dependent variable, we run similar model specifications as in the co-inventorship variable section above. We report the results from these regressions in Table 3.7 below. First, we look at the effect of the Chinese hs diaspora in the first and second DIOC edition – columns (1) to (3) and (4) to (6) respectively. Then we run the same regressions again for the case of the Indian hs diaspora in DIOC 00/01 – (7) to (9) – and DIOC 10/11 – (10) to (12). The baseline regressions for both first and second DIOC editions – (1) and (4) –, yield positive and statistically significant estimates for the coefficients of the Chinese hs diaspora variable, with a value of 0.292 and 0.221 respectively. This means, we find that if the probability of having two Chinese hs from a random draw of two individuals from the total hs population of two host countries doubles, this would lead to a 29.2% and a 22.1% increase in joint authorship between these countries in the periods covered by DIOC 00/01 and DIOC 10/11 respectively. In order to obtain the marginal effects, we apply the coefficient estimates to the mean values of our sample. In DIOC 00/01, this means an increase of the sample average co-authorship from 407 to 526 would come from doubling the sample average probability of having two Chinese hs in a country-pair random draw from 0.041 to 0.082. Similarly, in DIOC 10/11, doubling the sample average probability of having two Chinese hs – from 0.987 to 1.974 – would increase the sample mean co-authorship from 1,220 to 1,490. Moreover, these estimated coefficients remain positive and significant after controlling for the OHS migrant bilateral links $\ln\text{OHSBShare}_{ij}$ as shown in columns (2) and (5). In DIOC 00/01 the estimate modestly drops to 0.262, while in DIOC 10/11 that estimate is reduced by nearly half of the baseline value to 0.114. As for the coefficients for the OHS migrant bilateral links, we get positive and significant values of 0.566 and 0.406 in each DIOC edition as shown in columns (2) and (5) respectively. When removing all country pairs involving the US, the estimates for our main explanatory variable increase to 0.954 in DIOC 00/01 and almost remain unchanged in DIOC 10/11 with a value of 0.111 – see columns (3) and (6) respectively. We also obtain positive and significant estimates for the coefficient of the OHS migrant bilateral links in both DIOC editions.

As for the controls, the coefficients for the usual gravity covariates yield mitigated results in all of the first six columns. In general, we find positive and significant results of the coefficients for the colonial ties, common language and contiguity variables and a negative and significant estimate for the geographical distance variable in the baseline models. However, these effects disappear – except for the coefficient of geographical distance – in the models where we introduce the OHS migrant bilateral links variable $\ln\text{OHSBShare}_{ij}$. This

implies the effect of OHS migrant bilateral links on co-authorship suffices to absorb the effects of most of the common gravity covariates. Also, we find a negative and significant coefficient of the variable for the product of GDPs of -0.219 in columns from (4), suggesting countries joining their efforts in co-authorship activities do not always belong to the same income group. Additionally, we find countries level of publications to positively impact on joint authorship, as shown by the positive and significant coefficient for product of publication levels in two countries in all model specifications.

Findings on impact of Indian hs migrants on co-authorship are in line with those for the Chinese, but with minor exceptions. The results for our main variable coefficients $\ln\text{NOHSShare}_{ijk}$ do not differ much from results for the Chinese hs diaspora variable in the baseline model as shown in columns (7) and (11) for DIOC 00/01 and 10/11 respectively – 0.246 and 0.229. We find a 100 percentage point increase in the Indian hs migrant diaspora would result to an increase in co-authorship by 24.6% and 22.9% in the periods covered by first and second DIOC editions respectively. Similarly here, we use our sample average values to get the marginal effects. This means that, in DIOC 00/01, doubling the average probability of getting two Indian hs from a random draw from two host countries – from 0.071 to 0.142 – would lead to an increase in the average co-authorship from 407 to 507. In a similar fashion in DIOC 10/11, if we double that same average share of two Indian hs from 0.331 to 0.662, this would result in an increase in the sample average co-authorship size from 1220 to 1499. When introducing the OHS migrant bilateral links variable $\ln\text{OHSBShare}_{ij}$ into the baseline model, our regressions yield slightly weaker estimates for the Indian hs diaspora variable – 0.217 and 0.114 – in the first and second DIOC editions respectively – see columns (8) and (11). The estimates for the OHS migrant bilateral links remain positive and significant in all model specifications. Dropping the US from the regressions leads to an increase of the value of our main variable coefficients, 0.341 and 0.138 in DIOC 00/01 and 10/11 as it can be seen in columns (9) and (12) respectively. Other control variables and common gravity covariates estimates return results which are similar to what we had from the Chinese hs diaspora regressions insofar that these results are mitigated in their signs and significance, while we always get positive and significant estimates of the variable for the product of publications size in all columns from (7) to (12).

Table 3.7: Chinese and Indian hs diaspora in international co-authorship

Co-authorship	DIOC 00/01 CHINA			DIOC 10/11 CHINA			DIOC 00/01 INDIA			DIOC 10/11 INDIA		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
lnNOHSShare _{ijk}	0.292* (0.171)	0.262* (0.145)	0.954*** (0.164)	0.221*** (0.048)	0.114*** (0.040)	0.111*** (0.043)	0.246* (0.147)	0.217* (0.114)	0.341* (0.176)	0.229*** (0.060)	0.114* (0.072)	0.138** (0.065)
lnOHSBShare _{ij}		0.566*** (0.075)	0.407*** (0.089)		0.406*** (0.075)	0.292*** (0.104)		0.555*** (0.116)	0.366*** (0.133)		0.406*** (0.094)	0.274** (0.013)
Common lang.	0.129* (0.074)	0.062 (0.066)	0.111 (0.079)	0.159*** (0.058)	0.141*** (0.054)	0.225*** (0.078)	0.119 (0.088)	0.047 (0.081)	0.121 (0.082)	0.058 (0.071)	0.110 (0.073)	0.180** (0.074)
ln(distance)	-0.315*** (0.046)	-0.309*** (0.026)	-0.323*** (0.023)	-0.239*** (0.029)	-0.225*** (0.018)	-0.226*** (0.021)	-0.304*** (0.042)	-0.383*** (0.031)	-0.319*** (0.029)	-0.269*** (0.022)	-0.295*** (0.030)	-0.266*** (0.026)
Contiguity	0.150* (0.080)	-0.022 (0.070)	-0.061 (0.049)	0.179*** (0.059)	0.077 (0.053)	0.071 (0.051)	0.153* (0.078)	-0.102 (0.081)	0.055 (0.061)	0.140** (0.056)	-0.001 (0.060)	0.045 (0.054)
Colony	0.149* (0.088)	0.030 (0.071)	0.224*** (0.084)	0.108* (0.066)	0.013 (0.060)	0.113 (0.085)	0.060 (0.076)	0.058 (0.070)	0.084 (0.124)	0.159* (0.081)	0.084 (0.068)	0.081 (0.093)
ln(GDP _i *GDP _j)	-0.109 (0.140)	0.018 (0.069)	0.012 (0.066)	-0.219** (0.089)	-0.006 (0.070)	-0.044 (0.078)	0.170 (0.110)	-0.097* (0.052)	-0.134*** (0.052)	-0.246*** (0.068)	-0.340** (0.928)	-0.217*** (0.086)
ln(puli _i *publi _j)	0.890*** (0.039)	0.835*** (0.027)	0.761*** (0.039)	0.807*** (0.017)	0.765*** (0.018)	0.744*** (0.035)	0.817*** (0.042)	0.937*** (0.031)	0.752*** (0.029)	0.767*** (0.024)	0.869*** (0.025)	0.645*** (0.035)
Constant	-4.638* (2.560)	-6.293*** (1.297)	-5.194*** (1.329)	-1.323 (1.945)	-5.168*** (1.454)	-4.019*** (1.519)	-9.074*** (2.617)	-5.287*** (0.754)	-2.103*** (0.990)	0.204 (1.443)	0.553 (1.781)	1.481 (1.651)
Observations	465	465	435	465	465	435	465	465	435	465	465	435
Pseudo R2	0.973	0.982	0.979	0.976	0.983	0.977	0.972	0.975	0.974	0.976	0.975	0.977
Countries & time FE	Yes											
Without the US	No	No	Yes									

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

We add a unit to all of the above explanatory variables before logarithmic transformation in order to account of the presence of many zeros

3.5.2 Some robustness check

We check for the robustness of our results so far by means of a series of tests conducted in this section and in Appendix C. Here we test for whether our results might be driven by some generic effects from migration, regardless the skill level. Specifically, we test the hypothesis that the positive coefficients for our focal variables we obtained in the previous section could refer to the overall migration patterns, regardless of the skill level of migrants, as indeed is the case with trade. Additionally, this exercise allows us to discuss for potential endogeneity sources in our model – a parallel discussion of endogeneity is conducted in Appendix C with an instrumental variable analysis. In particular, one could argue that the controls we include in our model do not capture entirely the intensity of exchanges between each country pair, so that the estimated coefficients for our focal variables may be affected by a positive omitted variable bias. If this was the case, we would expect to get some positive effect not only from h_s migration, but also from all of the other skill groups, as the assumed hidden effect would likely affect the entire migration flows between these two countries and not just h_s .

Hence, we build variables for capturing links connecting migrants in each of the non-OECD low skilled – l_s – and medium skilled⁴¹ – m_s – diasporas – $NOLSShare_{ijk}$ and $NOMSShare_{ijk}$ respectively. We then proceed as in the previous section. That is, we calculate the product of the share of l_s or m_s Chinese/Indian population in the total l_s or m_s population of each country for pairs of countries. Additionally, we build OECD l_s and m_s migrant bilateral links variables – $OLSBShare_{ij}$ and $OMSBSHare_{ij}$ respectively. We thus run regressions for each of our two dependant variables – co-inventorship and co-authorship–, introducing these new skill variables into our initial equations. The results we get are illustrated in the Table 3.8 below.

For the co-inventorship regressions, the results show that controlling for Chinese or Indian l_s and m_s diasporas actually amplifies the effects of the h_s variable estimates as their magnitudes significantly increase as compared with what we had in Table 3.6 above. This increase applies to all model specifications except for column (6) where the estimate of the Indian h_s diaspora variable is smaller than the one we had in the baseline model. In contrast, we cannot find any significant effect of other skill group variables on co-inventorship. As for the OHS migrants’

⁴¹Low skilled are individuals having a primary education level only while medium skilled are those having a secondary education level.

bilateral links variable, its coefficients also increase considerably from the values we had in the baseline model in Table 3.6 above.

The co-authorship regressions in DIOC 00/01 return results from which can be drawn similar observations as with the co-inventorship ones: we get an increase of the estimates for the Chinese and Indian hs diaspora. However, this is not the case for the estimates in DIOC 10/11 whose effects actually diminish moderately. In parallel, we only get some positive and significant effect for the Indian ms diaspora estimates in DIOC 00/01 – see columns (5) and (6). The ls network estimates are negative in almost all columns.

Table 3.8: Other Chinese and Indian skill groups and international Co-inventorship and Co-authorship

	DIOC 00/01 CHINA		DIOC 10/11 CHINA		DIOC 00/01 INDIA		DIOC 10/11 INDIA	
Co-inventorship	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln\text{NOHSShare}_{ijk}$	2.111** (0.903)	1.627*** (0.606)	0.530*** (0.184)	0.398* (0.206)	1.077*** (0.300)	0.837*** (0.306)	0.548*** (0.157)	0.708*** (0.087)
$\ln\text{NOMSShare}_{ijk}$	-1.444 (4.408)	1.687 (2.950)	0.006 (0.260)	-0.121 (0.238)	1.109 (2.327)	-0.369 (2.320)	0.850 (0.816)	-1.118 (0.853)
$\ln\text{NOLSShare}_{ijk}$	-0.432 (0.922)	-0.541 (0.654)	-0.208 (0.146)	-0.195 (0.132)	-1.152 (1.153)	-0.546 (1.249)	-0.718 (0.522)	0.617 (0.567)
$\ln\text{OHSBShare}_{ij}$		1.467*** (0.515)		1.256*** (0.412)		1.695** (0.664)		0.934** (0.411)
$\ln\text{OMSBSHare}_{ij}$		-1.142 (1.195)		-1.362 (1.155)		-1.531 (1.774)		-0.341 (1.047)
$\ln\text{OLSBSHare}_{ij}$		0.299 (0.588)		0.424 (0.579)		-0.056 (0.882)		0.146 (0.568)
Co-authorship								
$\ln\text{NOHSShare}_{ijk}$	1.439*** (0.532)	1.301*** (0.352)	0.187* (0.110)	0.165* (0.0966)	0.269* (0.146)	0.241* (0.133)	0.168** (0.0741)	0.106* (0.0640)
$\ln\text{NOMSShare}_{ijk}$	-1.693 (1.725)	-1.606 (1.230)	-0.145 (0.146)	-0.0788 (0.137)	3.015*** (0.876)	1.952*** (0.588)	0.459 (0.281)	0.145 (0.202)
$\ln\text{NOLSShare}_{ijk}$	-1.751*** (0.532)	-1.616*** (0.399)	-0.154** (0.073)	-0.248*** (0.082)	-1.409*** (0.470)	-0.635** (0.309)	-0.190 (0.193)	0.0441 (0.166)
$\ln\text{OHSBShare}_{ij}$		0.591** (0.268)		0.481*** (0.151)		0.402* (0.241)		0.570*** (0.122)
$\ln\text{OMSBSHare}_{ij}$		0.080 (0.521)		-0.001 (0.334)		0.014 (0.454)		-0.226 (0.301)
$\ln\text{OLSBSHare}_{ij}$		-0.154 (0.272)		-0.163 (0.217)		-0.0230 (0.271)		-0.0226 (0.197)
Observations	465	465	465	465	465	465	465	465

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

N = 465 for all regressions. All regressions include the full list of covariates, countries and times fixed effect and a constant as shown in Table 3.6 and Table 3.7 for co-inventorship and co-authorship as dependent variables respectively. We add a unit to all of the above explanatory variables before logarithmic transformation in order to account of the presence of many zeros.

Another issue with our results concerns our focus of Indian and Chinese migrant, which was deliberately arbitrary and motivated only by the importance of these two diaspora groups. One may doubt of the interpretation we provide of our results, based as it is on a view of knowledge as tacit and of migrants' social networks as important knowledge carrier, to the

extent that, if true and relevant, such results ought to hold also for other diaspora groups. Hence, we replicate our baseline exercise with ten other NOHS diasporas, all of which belong to the top hs migrants sending countries to the 31 OECD countries of our sample (see the list in Table B6 in Appendix B).

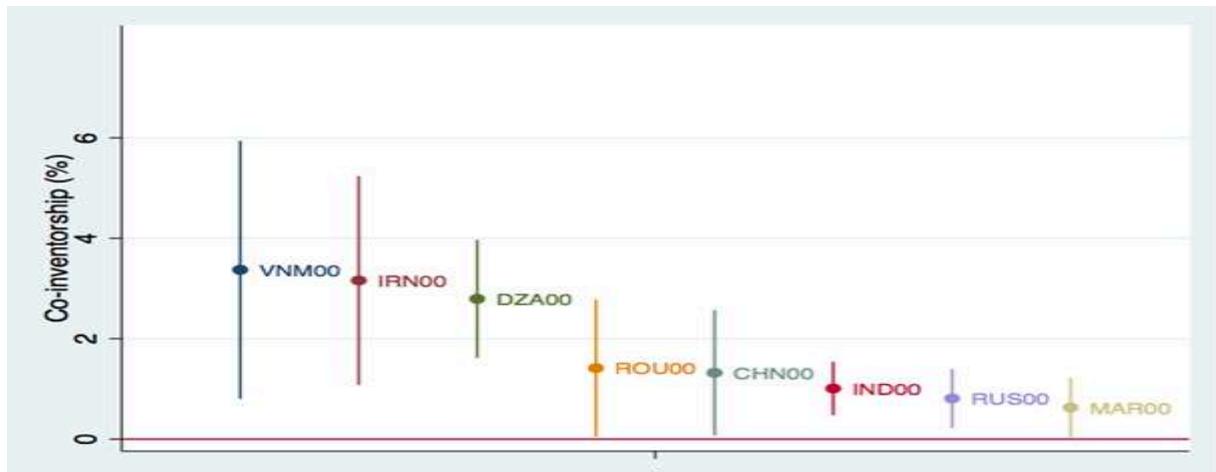
More specifically for each NOHS diaspora k among the ten, we compute the S&T collaborations elasticities with respect to the size of the product of NOHS migrant k shares in host countries i and j – NOHSShare_{ijk} . Following once more Felbermayr et al. (2010), we run separate regressions for each NOHS diaspora k and in each DIOC edition. Results from these regressions are summarized in Tables B4 and B5 Appendix B, for each S&T collaboration variable co-inventorship and co-authorship respectively. In what follows, we simply report estimated elasticities through a few graphs.

In general, we find that the positive results we got for the case of the Chinese and Indian hs diasporas extend to other NOHS diasporas, similar to what was found in the trade literature for overall migration (Felbermayr et al., 2010). Furthermore, for the diasporas shown in the below graphs, in most cases their effects are not significantly different one from another.

Figure 3.4 below reports the point estimates obtained for each NOHS diaspora variable from separate regressions – co-inventorship equation – as dots at the center of the spikes. The upper part represents DIOC 00/01 edition and the lower part DIOC 10/11. The spikes denote the 95% confidence intervals of each coefficient estimate – all shown estimates are statistically significant at least at the 1% level. The figure shows the Chinese and Indian hs migrants are not the most influential NOHS diasporas when it comes to co-inventorship. Indeed, the Vietnamese and Iranian hs diaspora effects on co-inventorship appear to be relatively higher in DIOC 00/01 and DIOC 10/11 respectively –3.372 and 3.159. Iranian and Pakistanis hs migrants are found at the second position in DIOC 00/01 and DIOC 10/11 respectively, while Chinese and Indian hs migrants only win a fifth and sixth position respectively in DIOC 00/01. The marginal effects of these NOHS migrant variables are presented in Table B3 in Appendix B. This table shows Algerian hs to have the highest marginal effect on co-inventorship in DIOC 00/01 and DIOC 10/11 as in average an increase of their share in country-pair destination – by 0.007 and 0.008 respectively - would lead to an increase co-inventorship instances by 268 and 63 respectively.

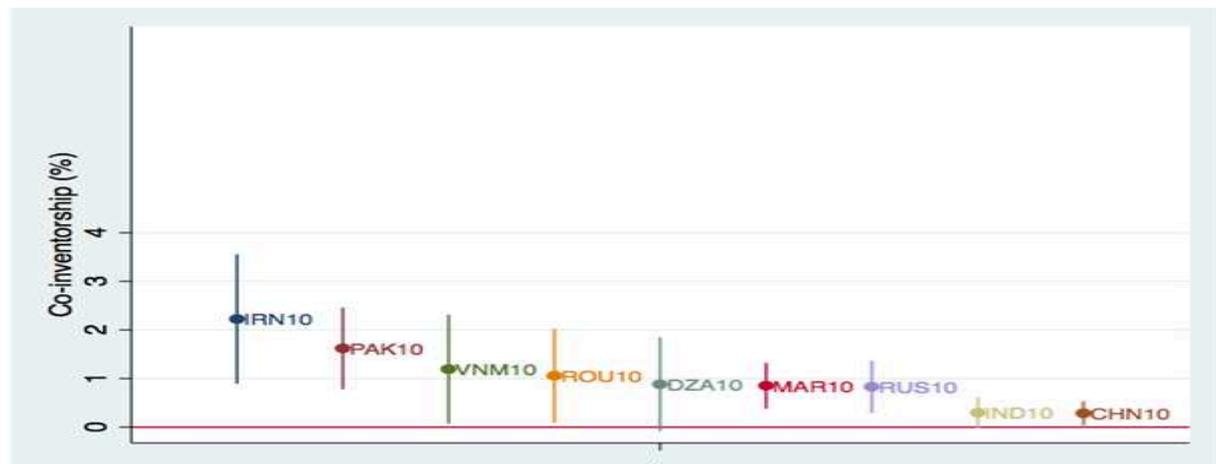
Figure 3.4: Co-inventorship elasticities of different NOHS diasporas

DIOC 00/01



VNM00 = Vietnamese hs diaspora DIOC 00/01 - IRN00 = Iranian hs diaspora DIOC 00/01 - DZA00 = Algerian hs diaspora DIOC 00/01 - ROU00 = Romanian hs diaspora DIOC 00/01 - CHN00 = Chinese hs diaspora DIOC 00/01 - IND00 = Indian hs diaspora DIOC 00/01 - RUS00 = Russian hs diaspora DIOC 00/01 - MAR00 = Moroccan hs diaspora DIOC 00/01 The dots represent the co-inventorship elasticities resulting from an increase in the NOHS diaspora k share by 1%.

DIOC 10/11

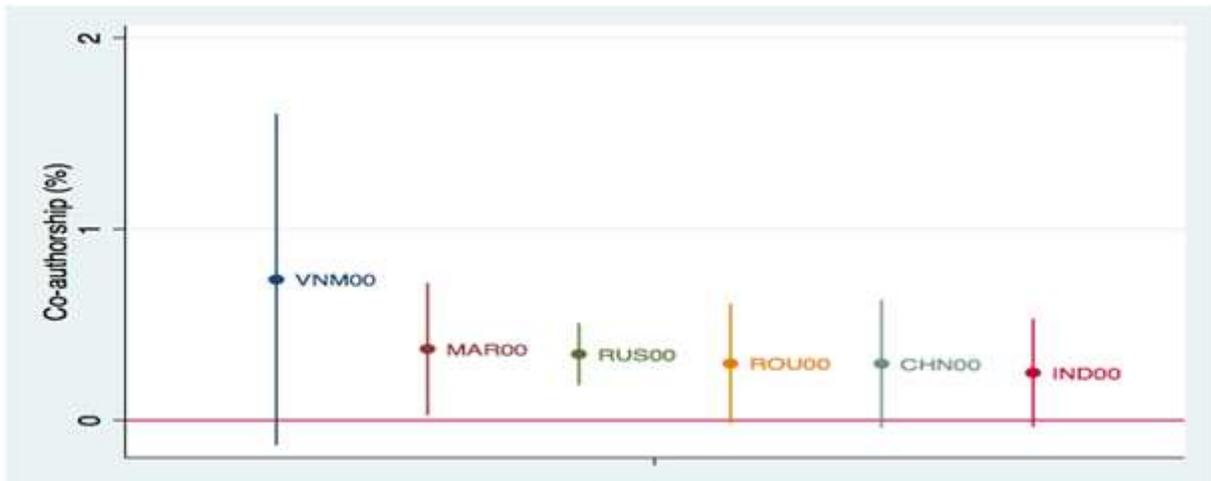


IRN10 = Iranian hs diaspora DIOC 10/11 -PAK10 = Pakistanis hs diaspora DIOC 10/11 -VNM10 = Vietnamese hs diaspora DIOC 10/11 -ROU10 = Romanian hs diaspora DIOC 10/11 - DZA10 = Algerian hs diaspora DIOC 10/11 - MAR10 = Moroccan hs diaspora DIOC 10/11 - RUS10 = Russian hs diaspora DIOC 10/11 -IND10 = Indian hs diaspora DIOC 10/11 -CHN10 = Chinese hs diaspora DIOC 10/11The dots represent the co-inventorship elasticities resulting from an increase in the NOHS diaspora k share by 1%.

The point estimates obtained for each NOHS diaspora from separate co-authorship regressions are shown in Figure 3.5 below. Similarly to co-inventorship regressions, all the reported estimates are statistically significant at least at the 1% level, with DIOC 00/01 edition in the top graph and DIOC 10/11 in the bottom one. Here again, the Chinese and Indian hs diasporas do not appear to have the biggest effect on co-authorship. Vietnamese hs diaspora induce the most important effect on co-authorship in DIOC 00/01 – 0.736 – while in DIOC 10/11 Pakistanis hs diaspora seem to entail the strongest effect– 0.653. Table B3 in Appendix B reports the marginal effect of Vietnamese hs diaspora on co-authorship in DIOC 00/01, which is, if their share in country-pair destination increases by 0.012, co-inventorship would increase by 300. While in DIOC 10/11, the marginal effect of Pakistanis hs diaspora denotes an increase of their share by 0.015 induces an increase in co-authorship by 797.

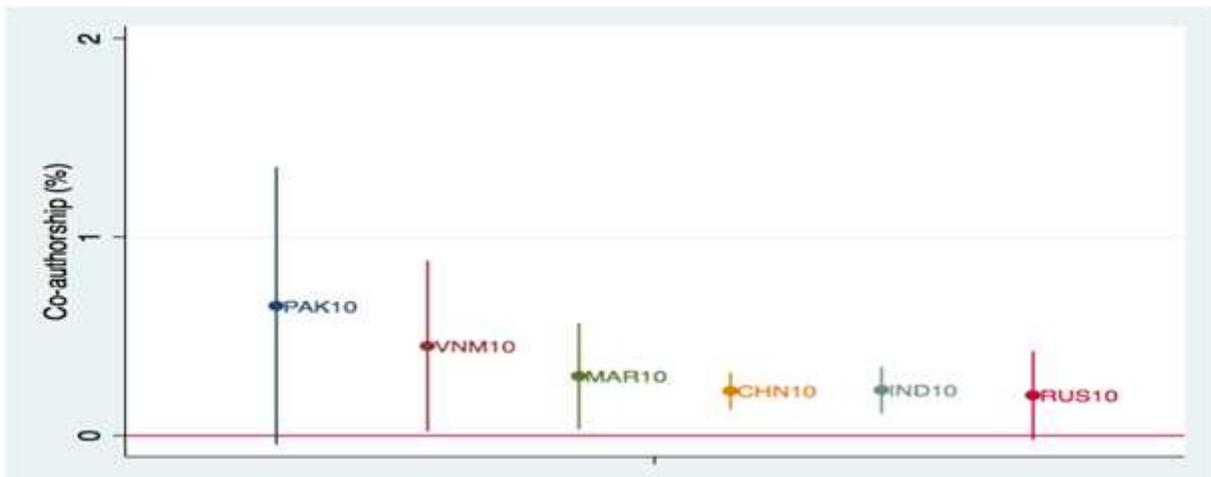
Figure 3.5: Co-authorship elasticities of different NOHS diasporas

DIOC 00/01



VNM00 = Vietnamese hs diaspora DIOC 00/01 - MAR00 = Moroccan hs diaspora DIOC 00/01 - RUS00 = Russian hs diaspora DIOC 00/01 -ROU00 = Romanian hs diaspora DIOC 00/01 -CHN00 = Chinese hs diaspora DIOC 00/01 -IND00 = Indian hs diaspora DIOC 00/01 The dots represent the co-authorship elasticities resulting from an increase in the NOHS diaspora k share by 1%.

DIOC 10/11



PAK10 = Pakistanis hs diaspora DIOC 10/11 -VNM10 = Vietnamese hs diaspora DIOC 10/11 -MAR10 = Moroccan hs diaspora DIOC 10/11 -CHN10 = Chinese hs diaspora DIOC 10/11 -IND10 = Indian hs diaspora DIOC 10/11 -RUS10 = Russian hs diaspora DIOC 10/11. The dots represent the co-authorship elasticities resulting from an increase in the NOHS diaspora k share by 1%.

3.6 Concluding remarks

This paper aims at filling an existing gap in the hs migration literature by analysing the relationship between hs diasporas within specific receiving countries and S&T collaboration between these countries. More precisely, we have investigated the role played by two of the most influential non-OECD migrant groups – Chinese and Indian – on S&T collaboration across OECD destination countries. We have assessed two types of S&T collaborations: co-inventorship and co-authorship. Based on a gravity model, we have undertaken a cross-section analysis at a country-pair level, with our main explanatory variables capturing the potential size of social networks within each specific hs diaspora group. We find that both Chinese and Indian hs diasporas have a strong and positive impact on all our two dependent variables in both DIOC editions. Our results hold also when controlling for OECD hs migrant bilateral links, namely for links between the hs migrants from each pair of destination countries considered in our sample. In addition we find similar, and sometimes stronger effects after controlling for migration at two further skill levels, low and middle. This enables us to dismiss the possibility that our results might in fact capture some generic effects of migration in general, instead of some specific effects linked to hs migrants. In a further analysis, we replicate this exercise for 10 of the most important non-OECD hs diasporas besides the Chinese and Indian ones. Our findings suggest that although the latter are the most influential in absolute terms, there are other hs diasporas with positive and significant effects on S&T collaboration. In particular, we found comparable effects for the Iranian, Pakistani and Vietnamese hs diasporas.

Overall, our results hold after several robustness tests, suggesting a causality effect of the presence of hs diasporas on S&T collaboration among host countries. These results point to the importance of maintaining and strengthening linkages within hs diasporas abroad or internationally across destination countries. These linkages have the potential to favor knowledge exchange between destination countries through similar mechanisms as the ones suggested by the trade literature, such as lowering transaction costs and reducing informal barriers.

However, since our analysis was limited to a reduced list of hs diasporas within specific host countries we should refrain ourselves to generalize our results. Therefore, one interesting extension to this paper would be to conduct a similar analysis for the case of hs diasporas from developing countries within non-OECD countries destination countries – like for

instance from Africa or developing countries in general. The proposed extension is relevant to the extent that South-South level of analysis is scant in this literature. Furthermore, it would be an important contribution to shift the attention from the '*brain drain*' narrative that has been at the core of the debate on South-North hs migration debate towards the '*brain gain*' potential of South-South hs migration.

Appendix B – Some tables

Table B1: Detailed migration data sources by country of destination.

DIOC 00/01		DIOC 10/11	
COUNTRY	SOURCE	COUNTRY	SOURCE
Australia	Census, 2001	Australia	Census, 2011
Austria	Census, 2001	Austria	European Labour Force Survey 2010/11
Belgium	ESEG, 2001	Belgium	Census, 2011
Canada	Census, 2001	Canada	National Household Survey (NHS) 2011
Chile	Census, 2002	Chile	The National Socio-Economic Survey, 2011
Czech Republic	Census, 2001	Czech Republic	Census, 2011
Denmark	Register, 2002	Denmark	Population Register 2011
Estonia	Census, 2000	Estonia	Census, 2011
Finland	Register, 12/2000	Finland	Population Register 2010
France	Census, 1999 LFS, 1998-2002, 2005	France	Census, 2011
Germany		Germany	Micro Census, 2011
Greece	Census, 2001	Greece	Census, 2011
Hungary	Census, 2001	Hungary	Census, 2011
Ireland	Census, 2002	Ireland	Census, 2011
Israel	LFS, 2001	Israel	Labour Force Survey 2011
Italy	Census, 2001	Italy	Census, 2011
Japan	Census, 2000	Japan	Census, 2010
Luxembourg	Census, 2001	Luxembourg	Census, 2011
Mexico	Census, 2000	Mexico	Census, 2010
Netherlands	LFS, 1998-2002	Netherlands	Census, 2011
New Zealand	Census, 2001	New Zealand	Census, 2013
Norway	Registers, 12/2003	Norway	Population Register 2011
Poland	Census, 2001	Poland	Census, 2011
Portugal	Census, 2001	Portugal	Census, 2011
Slovakia	Census, 2001	Slovakia	Census, 2011
Slovenia	Census, 2002	Slovenia	Census, 2011
Spain	Census, 2000	Spain	Census, 2011
Sweden	Registers, 12/2003	Sweden	Population Register 2010
Switzerland	Census, 2000	Switzerland	European Labour Force Survey 2010/11
United Kingdom	Census, 2001	United Kingdom	Census, 2011
USA	Census, 2000	USA	American Community Survey 2007-2011

Notes: ESEG: Enquête socio-économique générale; LFS: Labour force survey.

Sources : DIOC 2010/11 methodology & DIOC 2000/01 methodology.

Table B2: Descriptive statistics

Variable	DIOC 00/01					DIOC 10/11				
	N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max
OECD DESTINATION COUNTRIES										
Co-inventorship	465	96	403.495	0	4928	465	72	279.75	0	3388
Co-authorship	465	407	980.492	0	9254	465	1220	2377.304	2	22334
R&D cooperation	465	212	505.732	0	5028	406	158	368.168	0	3055
Hs bilateral shares	465	0.092	0.653	0	9.684	465	0.089	0.527	0	7.294
CHINA										
Hs shares	465	0.041	0.151	0	1.943	465	0.987	6.817	0	96.74
INDIA										
Hs shares	465	0.071	0.265	0	2.414	465	0.331	1.198	0	12.046
RUSSIA										
Hs shares	465	0.727	8.222	0	174.300	465	0.618	4.203	0	86.259
ROMANIA										
Hs shares	465	0.078	0.372	0	6.484	465	0.126	0.322	0	3.548
VIETNAM										
Hs shares	465	0.012	0.041	0	0.442	465	0.027	0.108	0	1.522
IRAN										
Hs shares	465	0.021	0.049	0	0.429	465	0.034	0.069	0	0.665
MOROCCO										
Hs shares	465	0.034	0.167	0	2.600	465	0.044	0.163	0	1.956
PAKISTAN										
Hs shares	465	0.004	0.015	0	0.207	465	0.015	0.051	0	0.626
ALGERIA										
Hs shares	465	0.007	0.041	0	0.445	465	0.008	0.039	0	0.478

Table B3: Marginal effects of co-inventorship and co-authorship

	China	India	Russia	Romania	Iran	Morocco	Pakistan	Vietnam	Algeria
DIOC 00/01									
ΔHs shares	0.041	0.071	0.727	0.078	0.021	0.034	0.004	0.012	0.007
ΔCo -inv.	127	97	78	136	303	60	-----	324	268
ΔCo -aut.	119	100	147	121	-----	152	-----	300	-----
DIOC 10/11									
ΔHs shares	0.987	0.331	0.618	0.126	0.034	0.044	0.015	0.027	0.008
ΔCo -inv.	21	22	60	76	160	61	117	86	63
ΔCo -aut.	270	279	245	-----	-----	365	797	550	-----

Table B4: Other NOHS diaspora co-inventorship elasticities

Co-inventorship	DIOC 00/01 RUSSIA						DIOC 10/11 RUSSIA						DIOC 00/01 ROMANIA			DIOC 10/11 ROMANIA			
	VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)						
lnNOHSShare _{ijk}	0.808*** (0.299)	0.626*** (0.235)	0.422*** (0.156)	0.830*** (0.274)	0.530*** (0.202)	0.605*** (0.227)		1.414** (0.695)	1.122* (0.672)	1.012* (0.609)	1.057** (0.493)	0.883** (0.446)	0.993** (0.465)						
lnOHSBShare _{ij}		0.948*** (0.182)	0.569*** (0.179)		0.776*** (0.139)	0.513** (0.224)			0.799*** (0.147)	0.661*** (0.179)		0.656*** (0.138)	0.562*** (0.214)						
Constant	-59.51*** (11.42)	-50.70*** (2.684)	9.127** (3.624)	-2.960 (2.294)	-2.775 (4.166)	7.908* (4.239)		-4.334 (3.626)	-8.539*** (3.061)	1.908 (4.869)	-10.52*** (2.873)	-34.86*** (3.349)	-5.532 (3.861)						
Observations	465	465	435	465	465	435		465	465	435	465	465	435						
Pseudo R2	0.964	0.981	0.970	0.989	0.994	0.982		0.986	0.990	0.969	0.987	0.992	0.978						
	DIOC 00/01 PHILIPPINES						DIOC 10/11 PHILIPPINES						DIOC 00/01 VIETNAM			DIOC 10/11 VIETNAM			
lnNOHSShare _{ijk}	-0.343 (0.373)	-0.413 (0.294)	0.138 (0.449)	-0.172 (0.188)	0.274 (0.167)	-0.037 (0.120)		3.372** (1.311)	3.428** (1.423)	2.533** (1.104)	1.192** (0.571)	0.530 (0.573)	-0.274 (0.469)						
lnOHSBShare _{ij}		0.847*** (0.151)	0.796*** (0.171)		0.649*** (0.155)	0.608** (0.208)			1.172*** (0.298)	0.669*** (0.164)		0.800*** (0.149)	0.676*** (0.210)						
Constant	20.26** (8.664)	1.075 (6.362)	-1.036 (5.509)	8.126 (25.70)	-64.67*** (24.73)	6.878 (9.152)		-51.82** (22.44)	-34.89 (26.04)	5.104 (4.572)	-61.50*** (19.35)	-67.14*** (20.13)	-0.650 (7.072)						
Observations	461	461	431	461	461	431		461	461	431	461	461	431						
Pseudo R2	0.987	0.991	0.967	0.986	0.991	0.976		0.934	0.948	0.969	0.987	0.993	0.973						
	DIOC 00/01 IRAN						DIOC 10/11 IRAN						DIOC 00/01 MOROCCO			DIOC 10/11 MOROCCO			
lnNOHSShare _{ijk}	3.159*** (1.062)	3.313*** (0.955)	2.568* (1.356)	2.224*** (0.679)	2.619*** (0.866)	1.749** (0.763)		0.628** (0.303)	0.466* (0.273)	0.519* (0.282)	0.851*** (0.240)	0.877*** (0.223)	0.786*** (0.231)						
lnOHSBShare _{ij}		0.841*** (0.152)	0.707*** (0.168)		0.566*** (0.177)	0.611*** (0.209)			0.778*** (0.141)	0.668*** (0.164)		0.373*** (0.127)	0.586*** (0.207)						

Constant	-49.72*** (11.50)	-51.73*** (10.16)	15.85*** (5.040)	-76.97*** (13.58)	-81.83*** (13.96)	12.95*** (3.760)	-5.557 (3.842)	-7.444** (3.194)	-0.041 (3.678)	-12.95*** (3.117)	-24.69*** (3.322)	-6.583*** (2.495)
Observations	464	464	434	464	464	434	463	463	433	463	463	433
Pseudo R2	0.985	0.990	0.969	0.987	0.990	0.977	0.987	0.990	0.971	0.990	0.988	0.976
DIOC 00/01 PAKISTAN			DIOC 10/11 PAKISTAN			DIOC 00/01 COLOMBIA			DIOC 10/11 COLOMBIA			
lnNOHSShare _{ijk}	-0.204 (1.436)	-1.227 (1.575)	-0.642 (1.295)	1.621*** (0.429)	0.896** (0.384)	1.041*** (0.383)	-3.414 (4.748)	-1.635 (4.524)	-12.74 (9.329)	3.013 (3.003)	-0.741 (2.459)	-0.927 (3.295)
lnOHSBShare _{ij}		0.827*** (0.148)	0.697*** (0.166)		0.730*** (0.148)	0.580*** (0.211)		0.596*** (0.190)	0.828*** (0.286)		0.203 (0.274)	1.004*** (0.359)
Constant	-30.31** (15.37)	-33.98** (14.27)	10.52*** (3.753)	-33.66* (20.31)	-34.13* (20.09)	7.602* (4.151)	-49.28*** (11.24)	-48.73*** (11.07)	9.249** (4.594)	-24.31* (13.50)	-38.79** (18.95)	-17.15 (11.72)
Observations	464	464	434	464	464	434	465	465	436	465	465	436
Pseudo R2	0.986	0.990	0.968	0.988	0.993	0.977	0.987	0.988	0.991	0.972	0.927	0.919
DIOC 00/01 CUBA			DIOC 10/11 CUBA			DIOC 00/01 ALGERIA			DIOC 10/11 ALGERIA			
lnNOHSShare _{ijk}	-2.569 (6.168)	-0.603 (6.007)	-4.509 (32.81)	-0.067 (4.307)	0.360 (4.008)	10.25 (15.18)	2.793*** (0.599)	2.435*** (0.536)	2.126*** (0.528)	0.881* (0.492)	0.459 (0.391)	0.347 (0.562)
lnOHSBShare _{ij}		0.620*** (0.176)	1.165*** (0.313)		0.746*** (0.185)	0.773* (0.443)		0.930*** (0.166)	0.725*** (0.166)		0.822*** (0.154)	0.663*** (0.210)
Constant	-74.60*** (25.40)	-72.92*** (25.35)	-3.747 (6.389)	-50.75 (51.26)	-48.88 (48.68)	-8.186 (8.360)	-100.4*** (13.13)	-99.67*** (11.06)	0.104 (4.298)	-9.685** (4.561)	-13.09** (5.642)	-4.794 (3.305)
Observations	465	465	435	465	465	435	465	465	435	465	465	435
Pseudo R2	0.988	0.989	0.974	0.964	0.973	0.949	0.972	0.981	0.971	0.988	0.993	0.976

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table B5: Other NOHS diaspora co-authorship elasticities

Co-authorship VARIABLES	DIOC 00/01 RUSSIA			DIOC 10/11 RUSSIA			DIOC 00/01 ROMANIA			DIOC 10/11 ROMANIA		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
lnNOHSShare _{ijk}	0.360*** (0.085)	0.282*** (0.076)	0.231*** (0.087)	0.201* (0.116)	0.166* (0.097)	0.192* (0.105)	0.297* (0.160)	0.240* (0.135)	0.362** (0.144)	0.136 (0.197)	0.0441 (0.170)	0.0602 (0.139)
lnOHSBShare _{ij}		0.520*** (0.080)	0.383*** (0.099)		0.449*** (0.075)	0.316*** (0.113)		0.576*** (0.098)	0.466*** (0.096)		0.473*** (0.076)	0.332*** (0.108)
Constant	-8.055*** (1.410)	-7.842*** (1.368)	-6.186*** (1.305)	-7.531*** (2.385)	-6.274** (2.625)	-2.403 (1.960)	-9.094*** (0.805)	-9.317*** (0.816)	-8.480*** (1.763)	-9.690*** (2.544)	-8.707*** (2.365)	-3.357** (1.374)
Observations	465	465	435	465	465	435	465	465	435	465	465	435
Pseudo R2	0.975	0.981	0.977	0.973	0.982	0.980	0.972	0.977	0.976	0.972	0.982	0.979
	DIOC 00/01 PHILIPPINES			DIOC 10/11 PHILIPPINES			DIOC 00/01 VIETNAM			DIOC 10/11 VIETNAM		
lnNOHSShare _{ijk}	-0.0318 (0.158)	-0.0758 (0.155)	0.572*** (0.153)	0.0574 (0.058)	-0.0176 (0.051)	-0.00563 (0.047)	0.736* (0.443)	0.806** (0.390)	1.584*** (0.400)	0.451** (0.220)	0.324* (0.190)	0.174 (0.223)
lnOHSBShare _{ij}		0.594*** (0.092)	0.475*** (0.092)		0.477*** (0.079)	0.398*** (0.097)		0.626*** (0.089)	0.507*** (0.093)		0.445*** (0.077)	0.342*** (0.107)
Constant	-9.783*** (2.630)	-10.30*** (2.625)	-7.663*** (2.139)	-10.34*** (2.282)	-9.017*** (2.124)	-5.312*** (1.567)	-8.265*** (0.834)	-6.859*** (0.788)	-7.480*** (2.183)	-9.768*** (2.376)	-5.314** (2.066)	-8.397*** (1.923)
Observations	461	461	431	461	461	431	461	461	431	461	461	431
Pseudo R2	0.973	0.980	0.975	0.970	0.980	0.976	0.959	0.973	0.976	0.970	0.983	0.979
	DIOC 00/01 IRAN			DIOC 10/11 IRAN			DIOC 00/01 MOROCCO			DIOC 10/11 MOROCCO		
lnNOHSShare _{ijk}	-0.124 (0.718)	-0.323 (0.671)	0.238 (0.501)	0.464 (0.420)	0.433 (0.383)	0.821** (0.387)	0.373** (0.176)	0.285* (0.154)	0.306** (0.145)	0.299** (0.135)	0.261** (0.129)	0.325** (0.137)
lnOHSBShare _{ij}		0.594*** (0.090)	0.489*** (0.096)		0.471*** (0.073)	0.406*** (0.097)		0.548*** (0.079)	0.419*** (0.103)		0.366*** (0.084)	0.249** (0.115)
Constant	-9.929***	-10.67***	-7.437***	-10.51***	-8.974***	-4.923***	-9.556***	-9.346***	-6.438***	-11.60***	-11.52***	-7.650***

	(2.621)	(2.682)	(2.177)	(2.285)	(2.161)	(1.623)	(0.764)	(0.677)	(0.865)	(1.873)	(1.805)	(1.226)
Observations	464	464	434	464	464	434	463	463	433	463	463	433
Pseudo R2	0.973	0.979	0.975	0.970	0.979	0.976	0.975	0.983	0.975	0.968	0.975	0.973
	DIOC 00/01 PAKISTAN			DIOC 10/11 PAKISTAN			DIOC 00/01 COLOMBIA			DIOC 10/11 COLOMBIA		
lnNOHSShare _{ijk}	0.356 (0.994)	-0.892 (1.204)	-0.844 (0.520)	0.653* (0.356)	0.160 (0.244)	-0.0436 (0.210)	-3.692 (3.519)	-1.548 (3.302)	6.128 (3.886)	-1.441 (1.134)	-1.506 (0.924)	-0.639 (0.789)
lnOHSBShare _{ij}		0.589*** (0.075)	0.460*** (0.091)		0.454*** (0.070)	0.329*** (0.101)		0.485*** (0.082)	0.399*** (0.102)		0.292*** (0.075)	0.183 (0.120)
Constant	-7.136*** (1.513)	-6.228*** (1.303)	-5.010*** (1.316)	-6.987*** (1.656)	-5.889*** (1.463)	-4.678*** (1.474)	-13.44*** (1.246)	-12.28*** (1.071)	-10.42*** (1.042)	-6.811*** (2.427)	-13.76*** (2.639)	-8.230*** (2.884)
Observations	464	464	434	464	464	434	465	465	436	465	465	436
Pseudo R2	0.975	0.983	0.978	0.974	0.983	0.977	0.980	0.983	0.989	0.982	0.984	0.982
	DIOC 00/01 CUBA			DIOC 10/11 CUBA			DIOC 00/01 ALGERIA			DIOC 10/11 ALGERIA		
lnNOHSShare _{ijk}	-4.172 (3.157)	-2.218 (3.029)	28.50*** (9.023)	-1.496 (0.984)	-0.878 (0.884)	14.63*** (4.790)	0.403 (0.440)	0.554 (0.471)	0.179 (0.353)	-0.0754 (0.362)	-0.0714 (0.392)	0.0555 (0.346)
lnOHSBShare _{ij}		0.439*** (0.095)	0.431*** (0.107)		0.271*** (0.089)	0.220* (0.128)		0.580*** (0.079)	0.482*** (0.101)		0.324*** (0.102)	0.388*** (0.123)
Constant	-11.10*** (1.690)	-10.59*** (1.741)	-10.93*** (1.332)	-11.57*** (2.502)	-10.64*** (2.414)	-12.02*** (2.637)	-4.515** (2.254)	-4.869** (2.112)	-4.811** (2.253)	-8.117*** (1.936)	-7.551*** (1.738)	-3.220** (1.409)
Observations	465	465	435	465	465	435	465	465	435	465	465	435
Pseudo R2	0.967	0.972	0.988	0.976	0.979	0.973	0.975	0.983	0.978	0.972	0.975	0.975

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table B6: Emigrant population 15+ in the OECD in 2010/11 by country origin

COUNTRY OF ORIGIN	ISO 3	Emigrant population ('000)	Highly educated emigrant population('000)
ALGERIA	DZA	1,504	306
CHINA	CHN	3,862	1,655
COLOMBIA	COL	1,217	365
CUBA	CUB	1,205	345
INDIA	IND	3,441	2,080
IRAN	IRN	845	424
MOROCCO	MAR	2,630	392
PAKISTAN	PAK	1,088	378
PHILIPPINES	PHL	2,854	1,417
ROMANIA	ROU	2,643	483
RUSSIA	RUS	1,953	660
VIETNAM	VNM	1,879	524

Source: DIOC 2010/11 <http://www.oecd.org/els/mig/dioc.htm>.

Appendix C – IV strategy

We examine the effect of Chinese and Indian hs migrants on co-inventorship and co-authorship after treatment for potential endogeneity through instrumental variable analysis. Drawing from Miguelez (2016), we use two instrumental variables: the size of the 1960 migrant group k in destination countries i and j (Özden et al., 2011) and the 1990 ls migrant group k in destination countries i and j (Docquier et al., 2009). We build the 1960 diaspora links for each specific migrant group k as the product of the 1960 migrant stocks within each destination country pair. The use of 1960 size is based upon two assumptions. On one hand, it is assumed that past immigration patterns to a specific destination strongly determine current immigration stocks and future flows in general. This in turn fuelled by network mechanisms, impacts on hs migration. While on the other hand, there should be no correlation between 1960 immigration stocks and current countries S&T collaboration as these immigration stocks record immigration flows from before the 1960s – a period of mass reconstruction within most OECD countries with a higher demand for less skilled labour. The second instrument is built as the product of the 1990 ls migrant group k stocks for each destination country pair. It represents a more contemporary variable given the period covered by our analysis. Again, here we assume past ls immigration to be determinant for current hs immigration stocks through networks operating within migrant communities, but also through training and education that ls migrants or their descendants might acquire at the destination country. On the contrary, there should be no correlation between past ls immigration and current countries S&T collaboration since in general there is no direct use of ls labour in S&T activities.

In Table C1 below, we show results from the GMM estimations of the PPML (Windmeijer & Santos Silva, 1997) for each dependent variable – co-inventorship and co-authorship – and for the Chinese and Indian hs migrants separately. A glance at the first column table values gives us an overview of the strength of our instruments. In columns (1) and (2), we see from the F-test statistics – which are 10.96 and 10.67 for DIOC editions 00/01 and 10/11 respectively – that the chosen instruments cannot be considered as weak. The Hansen J statistics for model specification test are reported at the bottom of columns (3) to (6). In all cases, our models seem not to exhibit any misspecification or over-identification issue. In general, the results suggest by omitting to account for endogeneity, we risk underestimating the real effect of hs migration on S&T collaboration as the estimates for each of the two hs diasporas considerably increase and remain significant as compared with results from the baseline models.

Table C1: GMM estimates with instrumented Chinese and Indian hs diasporas

CHINA	First-stage		GMM Co-inventorship		GMM Co-authorship	
	DIOC 00/01	DIOC 10/11	DIOC 00/01	DIOC 10/11	DIOC 00/01	DIOC 10/11
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
lnNOHSShare _{ijk}			2.489*** (0.796)	0.539*** (0.191)	0.402* (0.212)	0.313*** (0.102)
lnNOSize60 _{ijk}	0.004*** (0.001)	0.029*** (0.008)				
lnNOLSSize90 _{ijk}	0.003*** (0.001)	0.024*** (0.008)				
Constant	0.317*** (0.0495)	0.400** (0.203)	-19.05 (14.86)	-5.794 (4.691)	-0.761 (1.116)	-5.980*** (1.208)
F-test	10.14	20.18				
p-value	0.000	0.000				
Hansen's J chi2			1.608	10.062	9.191	9.760
p-value			0.658	0.122	0.163	0.135
INDIA						
lnNOHSShare _{ijk}			1.901*** (0.389)	0.653*** (0.140)	0.580*** (0.130)	0.449*** (0.096)
lnNOSize60 _{ijk}	0.018*** (0.003)	0.032*** (0.007)				
lnNOLSSize90 _{ijk}	0.005* (0.003)	0.014** (0.006)				
Constant	-0.274*** (0.074)	-0.492*** (0.152)	6.537* (3.561)	3.077 (2.805)	-8.950*** (1.005)	-5.266*** (1.156)
F-test	10.96	10.67				
p-value	0.000	0.000				
Hansen's J chi2			15.413	11.310	12.662	9.154
p-value			0.118	0.418	0.316	0.165

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.

We add a unit to all of the above explanatory variables before logarithmic transformation in order to account of the presence of many zeros. All regressions include the full list of covariates or controls, countries and times fixed effect and a constant. Total observations are N = 465.

Appendix D – Exploring R&D cooperation

In an extension of our paper, we investigate each of our focal NOHS diaspora effect on one important S&T collaboration variable which is R&D collaboration. Unlike co-inventorship and co-authorship which entail formal output from S&T collaboration, R&D cooperation captures the initial collaborative process leading to S&T production. Therefore, R&D cooperation stands more as an input measure. This thus makes R&D cooperation an imperfect or remote proxy of collaboration. However, R&D cooperation networks are broader to the extent they reflect at the same time basic and applied knowledge (Lata et al., 2012). Although, there remains a risk of capturing joint R&D efforts that will fail to result to an innovation or a scientific production. Yet the knowledge externalities to the involved partners, that occur during the collaboration process is non negligible. More importantly, the R&D cooperation variable points to collaborations decided at an organizational level. This point marks one of the key difference between this variable and the two other dependent variables from a collaboration incentive point of view – since co-inventorship and co-authorship are rather done at an individual level. Therefore, performing this exercise would help us getting a better understanding of the mechanisms behind findings from the co-inventorship and co-authorship analysis. That is, it will help us drawing a comparison between company or institution-related effects – from R&D cooperation – against individual attached effects – co-inventorship and co-authorship.

Our data source for R&D cooperation is the EU Framework Programme for Research and Technological Development (FP). Between 1984 and 2013 there have been seven FP waves, but due to our explanatory variable data constraint, we only consider the last three ones (FP5, FP6 and FP7). The FP database reports all R&D alliances and joint ventures that have been made under the auspice of the FP at an international level. Entities taking part of such alliances are individuals but mostly firms from the private sector, universities and other public institutions from the European Union (EU) and also from the rest of the world. To compute the R&D cooperation dependent variable, we proceed similarly as we did for the co-inventorship dependent variables. We perform the absolute counting of the number of co-partners in a joint venture, per country pair, per year as illustrated in the tables D1 and D2 below for a sample of two projects registered under the FP5:

Table D2: Sample of R&D projects under the FP5

Project id	Start Date	Contractor	
		Country	FP
		DE; ES; ES; FR;	
51424	01/02/00	IT; SE; SE	FP5
51426	01/02/00	DE; IL; IT; SE	FP5

Table D1: Counting of country pairs R&D cooperation

Ctry1	Ctry2	R&D		
		Year	cooperation	FP
DE	ES	2000	2	5
DE	FR	2000	1	5
DE	IT	2000	2	5
DE	SE	2000	3	5
DE	IL	2000	1	5
ES	FR	2000	2	5
ES	IT	2000	2	5
ES	SE	2000	4	5
FR	IT	2000	1	5
FR	SE	2000	2	5
IT	SE	2000	3	5
IL	IT	2000	1	5
IL	SE	2000	1	5

D.1 Chinese and Indian hs links impacts on international R&D cooperation

In Table D3 below, we present the results from the same model specifications as with the co-inventorship and co-authorship dependent variables for explaining R&D cooperation with Chinese and Indian hs diasporas alternatively.

The first column illustrates the results from the baseline model in DIOC 00/01 with the Chinese hs diaspora variable as the main explanatory variable. We find a significant estimate of 1.245 for this variable. This result means if we double the sample probability of getting two Chinese hs from a random draw of two individuals from the total hs population of two host countries, the size of their R&D cooperation would increase by of 124.5%. The marginal effects are computed using our sample average values for each of these variables. Therefore, an increase of the average country-pair Chinese hs share from 0.041 to a value of 0.082 would result to an increase in the R&D cooperation average value from 212 to the value of 476. Interestingly, we also find a similar positive effect of the estimate for the technological

similarity variable – a value of 1.227. In sharp contrast, the estimate for the product of R&D capacity in two countries does not give conclusive results as it is not significant. As for the common gravity covariates, we don't find any effect of their estimates except for the distance which is negative and significant. The second column shows the results from the baseline model regression with one additional variable denoting the OHS migrant bilateral links. As shown in column (2), the estimate for the Chinese hs diaspora just moderately increases to 1.300 and remains strong, while the estimate for the OHS migrant bilateral links is positive but not significant. However, the US seems to account for a major part of the effect of the Chinese hs diaspora in DIOC 00/01 as the strength of this variable estimate drops significantly when removing all observations with that country, as shown in column (3). Results from DIOC 10/11 only slightly differ from DIOC 00/01 results as described earlier. Indeed, from the baseline model in column (4) we find a strong estimate of 0.444 of the coefficient for the Chinese hs diaspora variable. That is, doubling the probability of getting two Chinese hs from a random draw of two individuals from two host countries would induce a raise in the country pair R&D cooperation by 44.4%. The marginal effect of this explanatory variable is derived from its sample mean value and the R&D cooperation one which are 0.987 and 138 respectively. So if this Chinese hs share doubles to 1.974, there would be a raise of R&D cooperation to 199. Interestingly, when adding the OHS migrant bilateral links variable in the baseline regression we get a significant estimate of 0.176 of this variable's coefficient, while the estimate for our main dependent variable nearly remains unchanged – see column (5). In columns (6) we see the US as a destination country accounts for all the effect of the OHS bilateral links variable as its coefficient loses its significance.

In contrast with the results we have for the case China, our results fail to find any impact of Indian hs diaspora on R&D cooperation in both first and second DIOC editions.

Table D3: Chinese and Indian hs diaspora in international R&D cooperation

R&D cooperation	DIOC 00/01 CHINA			DIOC 10/11 CHINA			DIOC 00/01 INDIA			DIOC 10/11 INDIA		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
lnNOHSShare _{ijk}	1.245*** (0.397)	1.300*** (0.395)	0.910* (0.537)	0.444*** (0.138)	0.422*** (0.133)	0.378*** (0.116)	-0.231 (0.175)	-0.254 (0.182)	-0.508 (0.344)	0.040 (0.182)	0.043 (0.185)	0.074 (0.090)
lnOHSBShare _{ij}		0.038 (0.062)	-0.014 (0.080)		0.176** (0.077)	-0.008 (0.172)		0.058 (0.060)	-0.123 (0.079)		-0.109 (0.185)	0.149* (0.078)
Common lang.	0.099 (0.077)	0.015 (0.051)	0.115 (0.101)	0.118 (0.104)	-0.025 (0.041)	0.147 (0.143)	0.028 (0.042)	0.008 (0.046)	0.338*** (0.108)	0.015 (0.101)	0.037 (0.108)	0.013 (0.040)
ln(distance)	-0.209*** (0.033)	-0.223*** (0.029)	-0.204*** (0.031)	-0.167*** (0.038)	-0.183*** (0.030)	-0.153*** (0.036)	-0.225*** (0.029)	-0.220*** (0.030)	-0.156*** (0.076)	-0.367*** (0.080)	-0.369*** (0.081)	-0.184*** (0.030)
Contiguity	0.024 (0.034)	0.025 (0.032)	0.030 (0.035)	0.009 (0.040)	0.006 (0.030)	0.011 (0.040)	0.028 (0.032)	0.024 (0.032)	-0.005 (0.090)	-0.141 (0.105)	-0.130 (0.105)	-0.002 (0.032)
Colony	-0.059 (0.060)	-0.033 (0.051)	-0.075 (0.056)	-0.106** (0.052)	-0.079* (0.044)	-0.067* (0.038)	0.003 (0.049)	-0.014 (0.051)	-0.259** (0.116)	-0.297* (0.167)	-0.302* (0.173)	-0.048 (0.047)
ln(RnD _i *RnD _j)	0.027 (0.026)	0.018 (0.028)	0.005 (0.016)	0.083* (0.048)	0.097** (0.043)	0.024 (0.033)	0.016 (0.025)	0.015 (0.025)	0.198*** (0.023)	0.102*** (0.038)	0.100*** (0.034)	0.062*** (0.016)
Tech. similarity	1.227*** (0.458)	0.780*** (0.243)	1.279*** (0.482)	0.772** (0.346)	0.278* (0.149)	0.754* (0.449)	0.877*** (0.250)	0.857*** (0.247)	7.286*** (0.730)	3.538*** (0.531)	3.600*** (0.541)	0.195 (0.156)
Constant	2.125*** (0.401)	2.297*** (0.333)	2.058*** (0.385)	1.890*** (0.345)	1.855*** (0.302)	1.785*** (0.341)	4.302*** (0.397)	4.267*** (0.398)	-0.206 (0.881)	4.219*** (0.976)	4.215*** (0.976)	4.537*** (0.415)
Observations	465	465	435	406	406	378	465	465	435	406	406	378
Pseudo R2	0.995	0.995	0.995	0.994	0.996	0.994	0.995	0.996	0.965	0.948	0.949	0.996
Countries & time FE	Yes											
Without the US	No	No	Yes									

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

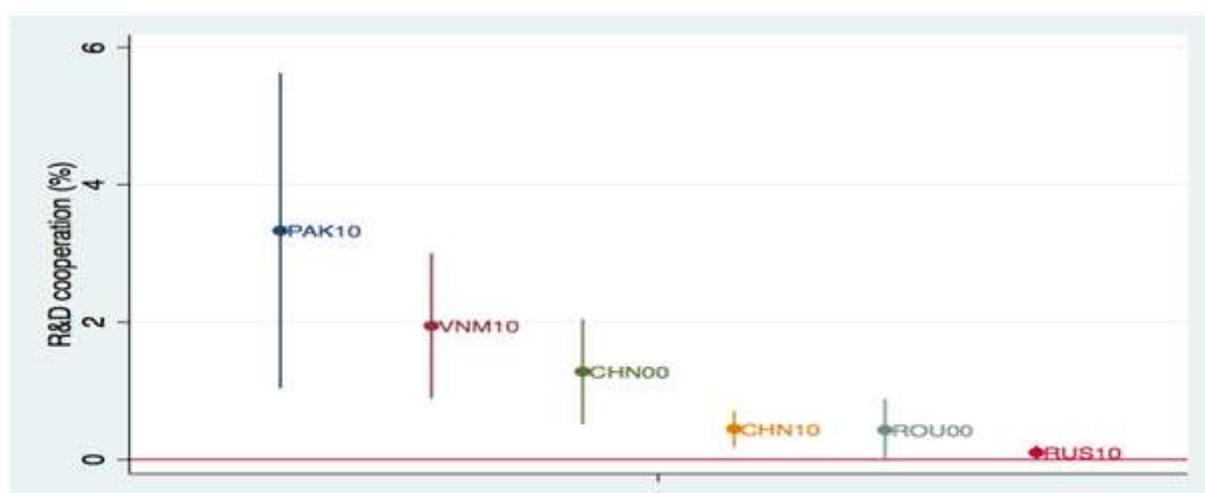
We add a unit to all of the above explanatory variables before logarithmic transformation in order to account of the presence of many zeros

D.2 Other NOHS diasporas R&D cooperation elasticities

We run regressions for our baseline model for the top 10 most important NOHS diasporas within OECD. In Figure D1 below, we present the point estimates from separate R&D cooperation regressions obtained for each NOHS diaspora variables. Since we only find statistically significant estimates at least at the 1% level for two NOHS diasporas in DIOC 00/01 and four NOHS diasporas in DIOC 10/11- see Table D4 below –, we show the two DIOC editions in a single graph. Those point estimates represent elasticity effects. Pakistanis hs migrants are shown as the most influential hs diasporas in DIOC 10/11 in terms of their effect on R&D cooperation – 3.332 –, followed by Vietnamese hs diaspora at the second position.

Figure D1: R&D cooperation elasticities of different NOHS diasporas

DIOC 00/01 and DIOC 10/11



PAK10 = Pakistanis hs diaspora DIOC 10/11 -VNM10 = Vietnamese hs diaspora DIOC 10/11 - CHN00 = Chinese hs diaspora DIOC 00/01 -CHN10 = Chinese hs diaspora DIOC 10/11 -ROU00 = Romanian hs diaspora DIOC 00/01 -RUS10 = Russian hs diaspora DIOC 10/11 The dots represent the R&D cooperation elasticities resulting from an increase in the NOHS diaspora k share by 1%.

Overall, we get positive significant results for only few cases out of the 12 initial NOHS diasporas we have investigated. These results point to the nature of this dependent variable which is more representative of incentives or decisions to collaborate taken at an institutional

level. Indeed, joint collaborations within the FP are likely to be initiated under an organizational setting and not by individuals. It is therefore unlikely that hs migrant exchanges which rather originate from hs migrants themselves, would have any significant effect on this variable.

Table D4: Other NOHS diaspora impact on R&D cooperation

R&D cooperation	DIOC 00/01 RUSSIA			DIOC 10/11 RUSSIA			DIOC 00/01 ROMANIA			DIOC 10/11 ROMANIA		
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\ln\text{NOHSShare}_{ijk}$	-0.270 (0.193)	-0.196 (0.148)	-0.0768 (0.115)	0.104** (0.0477)	0.0764* (0.0419)	0.0854* (0.0494)	0.429* (0.233)	0.666*** (0.238)	0.849*** (0.252)	0.323 (0.283)	0.365 (0.302)	1.293*** (0.271)
$\ln\text{OHSBShare}_{ij}$		-0.087 (0.063)	-0.066 (0.093)		0.146* (0.075)	0.139* (0.078)		-0.024 (0.095)	-0.113 (0.081)		-0.142 (0.204)	-0.004 (0.166)
Constant	3.467*** (0.884)	5.149*** (0.797)	3.729*** (0.575)	4.738*** (0.363)	4.749*** (0.362)	5.203*** (0.383)	3.647*** (0.779)	3.676*** (0.798)	0.0537 (0.871)	3.761*** (0.863)	3.731*** (0.861)	2.506*** (0.946)
Observations	465	465	435	406	406	378	465	465	435	406	406	378
Pseudo R2	0.987	0.992	0.993	0.996	0.996	0.996	0.957	0.963	0.973	0.947	0.948	0.959
	DIOC 00/01 PHILIPPINES			DIOC 10/11 PHILIPPINES			DIOC 00/01 VIETNAM			DIOC 10/11 VIETNAM		
$\ln\text{NOHSShare}_{ijk}$	-0.044 (0.382)	-0.086 (0.387)	-0.395 (0.417)	0.125 (0.133)	0.123 (0.132)	0.178 (0.110)	-0.280 (1.167)	-0.284 (1.086)	-0.359 (1.105)	1.944*** (0.539)	1.924*** (0.545)	1.635*** (0.360)
$\ln\text{OHSBShare}_{ij}$		0.043 (0.063)	0.062 (0.062)		0.184** (0.080)	0.160** (0.080)		-0.097 (0.062)	-0.058 (0.083)		-0.111 (0.113)	0.128 (0.080)
Constant	2.302*** (0.486)	2.278*** (0.487)	3.994*** (0.352)	3.189*** (0.454)	3.129*** (0.457)	3.858*** (0.349)	3.961*** (0.497)	4.519*** (0.652)	1.246 (0.967)	3.180*** (0.476)	3.176*** (0.474)	5.941*** (0.388)
Observations	461	461	431	402	402	374	461	461	431	402	402	374
Pseudo R2	0.995	0.995	0.996	0.996	0.996	0.996	0.992	0.992	0.979	0.992	0.992	0.996
	DIOC 00/01 IRAN			DIOC 10/11 IRAN			DIOC 00/01 MOROCCO			DIOC 10/11 MOROCCO		
$\ln\text{NOHSShare}_{ijk}$	0.669 (0.509)	0.667 (0.508)	0.798* (0.449)	0.428 (0.482)	0.422 (0.481)	0.720* (0.424)	0.123 (0.141)	0.121 (0.140)	0.128 (0.133)	-0.036 (0.091)	-0.041 (0.089)	-0.011 (0.084)
$\ln\text{OHSBShare}_{ij}$		-0.056 (0.087)	-0.062 (0.092)		0.058 (0.132)	0.025 (0.148)		-0.050 (0.083)	-0.058 (0.090)		0.084 (0.128)	0.043 (0.150)

Constant	4.364*** (0.940)	4.371*** (0.938)	3.351*** (0.549)	4.179*** (1.071)	4.185*** (1.072)	2.535*** (0.463)	4.723*** (0.889)	4.728*** (0.887)	3.587*** (0.521)	4.519*** (0.956)	4.532*** (0.958)	2.836*** (0.406)
Observations	464	464	434	405	405	377	463	463	433	404	404	376
Pseudo R2	0.992	0.992	0.993	0.994	0.994	0.994	0.992	0.992	0.992	0.994	0.994	0.994
	DIOC 00/01 PAKISTAN			DIOC 10/11 PAKISTAN			DIOC 00/01 COLOMBIA			DIOC 10/11 COLOMBIA		
lnNOHSShare _{ijk}	0.389 (1.083)	0.634 (1.122)	0.631 (1.163)	3.332*** (1.173)	3.234*** (1.125)	2.484** (1.094)	-0.851 (1.415)	-0.895 (1.417)	0.891 (1.413)	-0.583 (0.373)	-0.575 (0.376)	0.037 (0.753)
lnOHSBShare _{ij}		-0.036 (0.088)	-0.038 (0.091)		0.180** (0.080)	0.163** (0.081)		0.003 (0.065)	0.014 (0.067)		0.023 (0.096)	-0.246 (0.311)
Constant	4.237*** (0.586)	4.248*** (0.589)	6.145*** (0.530)	4.373*** (0.401)	4.369*** (0.405)	6.044*** (0.379)	2.704*** (0.341)	2.715*** (0.343)	2.309*** (0.398)	2.376*** (0.353)	2.369*** (0.356)	3.039*** (1.105)
Observations	464	464	434	405	405	377	465	465	436	406	406	379
Pseudo R2	0.994	0.994	0.994	0.996	0.996	0.996	0.996	0.996	0.996	0.997	0.997	0.950
	DIOC 00/01 CUBA			DIOC 10/11 CUBA			DIOC 00/01 ALGERIA			DIOC 10/11 ALGERIA		
lnNOHSShare _{ijk}	-5.756 (4.099)	-5.403 (4.086)	-5.081 (6.855)	-0.546 (2.263)	-0.004 (2.008)	2.950 (2.443)	0.120 (0.374)	0.108 (0.376)	0.096 (0.381)	0.024 (0.320)	-0.169 (0.415)	-0.045 (0.326)
lnOHSBShare _{ij}		0.040 (0.129)	0.065 (0.133)		0.212 (0.174)	0.200 (0.179)		-0.020 (0.078)	-0.021 (0.081)		0.081 (0.127)	0.165** (0.083)
Constant	3.116*** (0.934)	3.115*** (0.936)	3.131*** (1.070)	2.005*** (0.673)	2.404*** (0.421)	2.401*** (0.509)	4.203*** (0.584)	4.210*** (0.589)	5.875*** (0.531)	3.765*** (0.439)	4.948*** (0.976)	5.231*** (0.380)
Observations	465	465	435	406	406	380	465	465	435	406	406	378
Pseudo R2	0.972	0.972	0.972	0.958	0.982	0.982	0.994	0.994	0.993	0.995	0.994	0.996

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Chapter 4

Return migration and citations recency: the case of South African researchers

4.1 Introduction

Over the past quarter century, highly-skilled (hs) mobility has been one of the most dynamic component of international migration (Jonkers, 2008; OECD, 2008). Scientists have contributed extensively to what has become known as the ‘*global talent flow*’ (Kerr et al., 2016). Among other research topics, one of the most prominent is the contribution of such scientists to knowledge production in their home countries – ‘*brain gain*’. Indeed, the amount of qualitative and quantitative research on this topic has been growing consistently over the past two decades. This is in contrast with the ‘*brain drain*’ view that dominated the migration literature up to the 1990s, and mostly described hs migration as a pure loss for the home countries, poor ones in particular (Bhagwati & Hamada, 1974; Grubel & Scott, 1966). Data development has been crucial in fuelling this new trend of literature. However, most data mining efforts have targeted migration from Southern and Eastern Asian countries, particularly China and India, due to the sheer size of population shifts (Agrawal et al., 2011; Almeida et al., 2010; Chacko, 2007; Kerr, 2008). This paper targets instead Africa, which has so far received much less attention. In particular, it focuses on the Republic of South Africa (RSA or, simply, South Africa, thereon).

The ‘*brain gain*’ literature has identified several knowledge diffusion channels, one of which consists of ongoing social interactions between return migrants and their social networks in both the countries of destination and origin. This is relevant to the extent that even in science and technology, tacit knowledge needs to be transferred, and this requires physical and/or social proximity such as in face-to-face meetings or frequent exchanges (Stephan, 2010). A large literature – both within and outside migration – has tapped in bibliometric data sources and proved this point time and time again (Scellato et al., 2015; Trippel, 2013). In general, the idea of return migrants as an instrument of knowledge diffusion to their home country lies on

the assumption that these returnees have acquired some new knowledge from their stay abroad; knowledge that may have not been available to them had they not migrated. Moreover, it is assumed their foreign experience has made them exposed to the type of knowledge that may not be accessible to non-migrants; hence their likelihood to perform better than non-migrants upon their return (Kenney et al., 2013; Meyer, 2001; Wadhwa et al., 2009a, 2009b). The latter assumption is supported by empirical evidences on returnee scientists' scientific activities, particularly with their scientific productivity, international collaboration (Jonkers & Cruz-Castro, 2013; Jonkers & Tijssen, 2008; Scellato et al., 2015) in their home country. Some authors have argued that chances of returnees to succeed in their home country are greater when they maintain linkages with their former professional network in their former host country (Jin et al., 2007); while others have stressed on the key importance of the absorptive capacity of the home country (Song, 1997). Besides, returnees' academic training abroad – PhD. in particular – may positively affect their research activity home, not only from their accumulated capital benefits; but also from their integration into a transnational professional network of colleagues and supervisors from the foreign country where they have earned their degree (Shin et al. , 2014).. There has been an increasing flow of the most advanced knowledge for academic research within internationally-oriented knowledge networks (Mahroum, 2005) and as part of such networks, returnees the chance to absorb it.

We propose to investigate returnees' research performances by analysing the backward citations of publications by RSA returnee scientists – RSA scientists who have spent at least one year working abroad. In particular we are interested into these citations' '*recency*'⁴², to be intended as the closeness between the publication's year of a scientific article and the average year of the literature it cites (Buttler, 1999; Merton, 1968; Nkiko & Adetoro, 2007; Tagliacozzo, 1977). We use two recency indicators: i) the average age of citations (which has an inverse relationship with recency) and ii) the share of less than 3-year old citations in the citing paper's reference list. Our analysis relies on the assumption that returnees' past and/or current interactions with their international social and scientific network makes them more likely to be exposed to the most recent literature in their respective field.

⁴² This term is not to be confused with the '*recency*' notion from the cognitive psychology field of studies whereis it referred to as either the time when the last event occurred (Fader et al. , 2005) or the propensity to recall the latest event from a memory (Bjork & Whitten, 1974).

The bulk of the literature on scientists' return migration, when founded on bibliometric data, has addressed the relationship between returnees' past migration experience and their international research collaborations (mainly with their former host countries), scientific productivity, and research output quality (De Filippo et al., 2009; Decramer et al., 2013; Jonkers & Cruz-Castro, 2013; Jonkers, 2009; Roos et al., 2014; Scellato et al., 2015). Yet, there is still some grey area on this topic and in particular some shortage of evidence connecting return migration to citations pattern. This is mainly due to the lack of comprehensive data resulting from linking full bibliometric information and researchers' CVs. To the best of our knowledge, there are no previous studies that have analysed the relationship between past migration experience and citation recency. We intend to fill this gap in the literature, but also to contribute to remedy the scarcity of studies on return migration and knowledge diffusion in an African country setting. We also aim at contributing to the policy-oriented discussion on the role of return migrants in the development of their home countries. As such, this paper' topic belongs to a new research area which is at the crossroads of development, migration and innovation economics. Therefore, the originality of our approach resides in connecting two areas of the literature that have not been jointly addressed before.

We make use of rich dataset made available by the National Research Foundation (NRF) of South African, the main public funding agency for science (Barnard et al., 2016). We link these data to bibliometric information from Web of Science (WoS), in order to retrieve information on the South African scientists' publication and citation patterns, based on fuzzy matching of the scientists' names and surnames. This allows both to identify migrant and return migrant scientists, as well as non-migrant ones, and their publications. We then proceed to test the hypothesis by which return migrants' citation recency would be higher than non-migrants'. To address concerns about unobserved heterogeneity, we exploit some unique features of our data for a panel data analysis and undertake two different econometric exercises. First, we use a random model regression with years fixed effect on all return migrants and non-migrants in our dataset. Second, we perform a propensity score matching (PSM) analysis that allows us to obtain a one-to-one match of a smaller sample of RSA returnees with non-migrants with similar pre-migration characteristics. We then run conditional difference-in-difference (CDID) estimations on this reduced matched sample. The results of both exercises confirm our hypothesis.

In what follows, we present our literature review in section 2 and our methodology and data in section 3. We then proceed to illustrate our baseline results in section 4 and the results of our robustness check in section 5. Section 6 concludes.

4.2 Literature review

A longstanding tradition in migration economics has focussed on assessing positive returns from migration for origin countries. Based on household surveys, early studies mostly investigated the potential role of financial remittances in capital formation. However, with the growth of his migration, the attention has shifted to migrants' "knowledge remittances", and their potential contribution to knowledge production and innovation (Breschi et al., 2016). In particular, returnees involved in scientific, engineering or entrepreneurial activities are often found to be important carriers of knowledge otherwise unavailable in their home countries (Jonkers & Cruz-Castro, 2013; Kahn & MacGarvie, 2011; Kenney et al., 2013; Meyer, 2001; Wadhwa et al., 2009a, 2009b). In what follows, we first provide a definition of return migration, and then focus on the evidence concerning returnee scientists.

Defining return migration is not straightforward, due to both conceptual and empirical problems. For instance, Dustmann & Weiss (2007) define return migrants as migrants who settle back in their home country by their own choice, after having spent several years abroad. However, in practice, it is rather difficult to apply such a precise definition. For example, a number of studies based on survey data consider as migrants all individuals having spent at least a year abroad, which makes it difficult to distinguish between return and temporary migrants (Jonkers & Cruz-Castro, 2013; Velema, 2012) – a more detailed discussion is provided in Appendix F. Other studies have relied instead on information on special temporary immigration programs, whose beneficiaries were required to leave the host country after completing a predetermined set of activities, so that their return decision cannot always be considered as voluntary (Gibson & McKenzie, 2014; Kahn & MacGarvie, 2011).

At a more theoretical level, one may encounter further difficulties in distinguishing return migration from circular migration. With the latter, there are repeated migration experiences between home and host countries, involving repeated migration and return movements (Hugo, 2013). Migrant graduate students are a critical component of circular migration.

Indeed, students immigration in most receiving countries is often not perceived as a stand-alone immigration category, due to the constraints of limited length of stay or difficult visa

renewal conditions at the end of their study (Kahn & MacGarvie, 2011). However, many returnee students falling under such requirements tend to go back to their former host country – or to another destination – after some time spent home, for work or other short/medium term assignments. This relatively short/medium term mobility of returnee students across their home country and other destinations subsequently creates a dynamic of circular migration in their career path. There is an increasing number of immigrant students who return home upon study completion or after some years of experience in their host country (Finn, 2010); among which a sizeable share a voluntary return.

In our paper, we embrace, rather than reject, the complexity of aspects surrounding return migration and adopt a broader definition, one that encompasses return from temporary migration – over a year – both for education and work purposes.

In general, however, return migration remains an under-explored topic. This is largely due to data unavailability. In many countries, it is quite difficult to get figures on immigrant outflows, since immigration authorities rarely record this type of information. Hence, most studies have relied on smaller scale data such as surveys data (Dustmann & Görlach, 2016). This comes with the limitation of reducing the scale of analysis and its findings to specific groups of returnees. One of the most common methods for capturing the scientific contribution of scientist returnees has been to link these survey data to bibliometric data. The latter has provided a rich data source for a wide range empirical studies on knowledge diffusion – as measured by citations –, on research output and productivity – publications count – and on scientific collaboration – co-publication.

Coming to returnee scientists, they may contribute to knowledge transfer first and foremost through their direct exposure to a wider scientific community, which is expected to make them more productive than the non-migrant scientists. This is confirmed by empirical studies which show that scientist returnees have a higher productivity in terms of scientific publications and are involved in more international cooperation than non-migrants (De Filippo et al., 2009; Decramer et al., 2013; Jonkers & Tijssen, 2008; Jonkers, 2009; Roos et al., 2014; Scellato et al., 2015).

In addition, returnees can foster knowledge networks between their home countries and their previous locations. Indeed, some studies find that return migration is positively associated to scientific or technological collaboration between the home and former host countries (Jonkers & Cruz-Castro, 2013; Kahn & MacGarvie, 2011).

A growing body of literature, in particular, has focused on the specific role of migrant scientists on the development of the national science system in their home countries. For example, Jonkers & Cruz-Castro (2013), who study the scientific publications patterns of Argentinean returnees, find them to have a higher propensity than natives both to publish in journals with high-impact factors and to co-publish internationally and with their former foreign institution. More generally, returnee scientists may serve as a bridge not only between their former host and home countries, but also with the global scientific community (Velema, 2012). By doing so, they help their home countries to move closer to the knowledge frontier.

As noted by Zhou & Leydesdorff (2006), the returnees have been at the core of the emergence of home country science in countries like China, through an increase of international cooperation initiatives. This is also true for many currently developed countries such as Germany, whose reconstruction process and catching up with the global community after World War II required the active participation of their returnees in all national sectors and in the science sector in particular (Jöns, 2009).

Finally, the impact of returnees – hs in particular – might expand far beyond what is captured by publication – or patent data. Indeed, the presence of these returnees in their home countries may yield larger benefit of other nature. For instance, their teaching, mentoring, advising or entrepreneurship activities may induce some knowledge spillover to the locals. Therefore, in an analysis of the impact of returnees on their home country one needs to keep in mind those hidden externalities whose effects may be greater than what can actually be observed with the data in hand. Yet, there remains some grey areas in the observable outcomes that might result from returnees' activities – scientific activities – in their home country. We propose to focus on one specific aspect of these scientific activities, namely returnees' publication, by assessing the recency of references cited in these publications. We investigate the case of RSA returnees. Citations recency, which can be used as a signal to the future impact of new ideas (Mukherjee et al., 2017), remains an unexplored topic. Furthermore, this topic connection with migration analysis provides a brand new and challenging approach theoretically as much as empirically. Besides, our analysis is centred on an African country which is often an understudied continent in the hs migration literature.

4.3 Methodology and data

4.3.1 Specifications

We want to assess whether returnee RSA scientists build their research upon more recent or up-to-date literature than their colleagues with no experience abroad. We do so by testing whether their publications exhibit higher recency of backward citations— items in the publications’ reference list.

Our data are organized as an (unbalanced) panel in which T is the total length of researcher i’s career – number of years since his/her first work appointment (this includes PhD), but we choose to restrict our analysis to the period 1985-2014. For each year t we observe all the publications published by i, and their reference lists. Based on the publication year of such references, we can calculate their recency, according to various indicators (see below).

We formulate our baseline model as follows:

$$R_{it} = \alpha_0 + \alpha_1 Mig_{it} + \beta_n X_i + \gamma_m Z_{it} + \delta_t + \varepsilon_{it} \quad (4.1)$$

Where:

- R_{it} denotes the average recency of references in publications by researcher i at time t
- Mig_{it} is a migration time dichotomous variable. It takes value 1 if researcher i had a work experience abroad in years preceding t and 0 otherwise – we do not include years during which the researcher was abroad.
- X_i is the set of time-invariant variables capturing inherent individual characteristics.
- Z_{it} stands for the set of time-varying covariates.
- δ_t is the year fixed effect.
- ε_{it} is the error term.

For estimation, we adopt random effect model with robust standard-errors, as we want to preserve both our time-invariant and time-variant variables.⁴³

⁴³The results from the Hausman specification test suggest there is correlation between individual effects and other regressors. However, when we use a fixed effects model all time-invariant variables such as gender,

In a variation of the baseline specification, we add to model (4.1) a dummy variable taking value 1 for the years after the year of Master degree obtained abroad and 0 otherwise. This variable identifies researchers who got their education out of RSA.

The baseline specification is essentially descriptive. In fact, it does not account for the potential endogeneity that might occur from hidden factors which may determine migration. These hidden factors might be either individual characteristics or some characteristics of the scientist's university of affiliation first as a student and then as researcher. In order to deal with it, we experiment with a second specification, based upon conditional difference-in-difference (CDID) analysis of a reduced sample of treated and matched RSA researchers, which we obtain through Propensity Score Matching (PSM) (Abadie, 2005).

In the CDID analysis the migration-and-return experience is our treatment, of which we estimate the effect on the recency of citation references, conditional on the effects of other covariates that may have affected the treatment administration. We add two variables to the baseline model in equation (4.1): first, a dummy to account for the treatment and control groups; and, second, the interaction between the treatment and a period variable (before and after treatment). The CDID model is therefore formulated as follows:

$$R_{it} = \alpha_0 + \alpha_1 PMig_{it} + \alpha_2 Treatment_i + \alpha_3 PMig_{it} * Treatment_i + \beta_n X_i + \gamma_m Z_{it} + \delta_t + \varepsilon_{it} \quad (4.2)$$

Where

- $PMig_{it}$ represents our migration time dichotomous variable. It captures post-migration period – or post-pseudo migration for individual belonging to the control group.
- $Treatment_i$ is the dummy for the treatment group; that is, it indicates individuals that belong to the group of returnees.
- $PMig_{it} * Treatment_i$ is the interaction term. This interaction term stands as the CDID variable.

Since Ashenfelter & Card (1985), the difference-in-differences methods have been widely used to investigate the effects of discrete events such as policy changes or migration shocks.

ethnicity and field of research are dropped from the regression. It is therefore not possible to conduct an analysis on these variables, although the estimated coefficient for our main explanatory variable is similar to the one from the random effects in terms of magnitude and significance when running a fixed effects model with robust standard-errors. The model is clustered at the level of individual-year, allowing the residuals to be independent across individuals and years.

The basic difference-in-differences method originally applies to a situation where two observed groups are compared on the basis of their outcomes (Imbens & Wooldridge, 2007). There are two periods of observations for each of the groups. However, only one of the groups is subject to treatment during the second period. One of the main requirements of the difference-in-differences estimator is the assumption of parallel paths followed by the average outcomes for the treated and control groups in the absence of the treatment. Hence, the necessity to select a control group as close as possible to the treated group in their observed features.

The PSM helps in this regards; that is, it uses a score built from the distribution of observed covariates to sample down an original larger sample of potential controls to a smaller sample size of selected units whose scores match the treatment ones (Rosenbaum & Rubin, 1985). This results to a matched group (control) that exhibits similar features to the treated group before the treatment. In our case, the control group should include individuals who would have likely migrated at the time of their matched individual's first migration experience. Therefore, the migration potential of the control group might help us to considerably reduce the endogeneity issue not accounted for in equation (4.1), in which we include all RSA researchers of our larger sample.

4.3.2 Data

The data for our analysis come from two sources, namely the CVs of South African researchers collected by the National Research Foundation (NRF dataset) and the Web of Science (WoS), maintained by Clarivate Analytics. The NRF dataset contains most of the information we use, including each researchers' publications, from his/her earliest one (after 1975) to 2014. The raw NRF data amounts to over 4,000 researchers and 214,242 publications. WoS provides some key complementary information for such publications, in particular their reference lists, completed with the publication years of each cited source.

4.3.2.1 Publications

The National Research Foundation is a RSA governmental agency, whose main mission is to promote scientific research and to support the development of the national research capacity. Roughly every four years, all the researchers who consider the possibility to apply

for public funding must require to get rated by this agency. They provide their entire CV with various information, ranging from their biographical essentials to their education, professional history and publication lists. All these information are refereed by half a dozen international experts who draft reports on the basis of which the NFR expert committee will assign a rating to the researchers – roughly: A, B, C; for a full description of the rating system see Barnard et al. (2016). Hence, researchers have a high incentive to list all their research output. These range from peer-reviewed journal publications to publications in the public press, but also books, conference presentations, patents, and policy or technical reports. In the present analysis, we focus on peer-reviewed journal publications, as they represent the highest-value output from academic research⁴⁴. Unfortunately, the publication data one can find in the NFR dataset, being the result of self-reporting, are not standardized – for example, the journal titles may be altered, shortened in a causal way or even missing – and miss out some important information, such as the authors’ affiliations and the reference lists. Filling these gaps requires an extensive work of data collection and cleaning.

Moreover, the individual biographies may be inaccurate or incomplete. In particular, the names of the institutions where the researchers obtained the education or worked may be misreported or lack indication of graduation or tenure years. Most crucially, in most of these records, the information on the country where these institutions are located is missing, which leaves us with no choice but to inferring it from the institutions’ names and a related geo-localization effort.

The NFR data come in two editions. The oldest one records applications filed between 2002 and 2012 – used in Barnard et al. (2016) – ; while the most recent one has application records spanning from 2012 to 2014. Combining both editions results into richer and more comprehensive data, in terms of size and longitudinal dimension of our sample. However, this comes with one major constraint, as RSA researchers’ identifiers are not harmonized throughout both editions. That is, a single researcher is assigned a different identifier in each data edition. Additionally, there is a lack of institution names harmonization throughout each data edition and across both of them.

⁴⁴ Peer-reviewing is a process through which a researcher’s work is critically assessed by his/her peers within the same scientific field before publication. Therefore, peer-reviewed journals are generally assumed to be a guarantee and standard of quality, against which all other types of research are judged (Roessner, 2002; Weingart, 2005).

Summing up, the raw NRF data require an extensive mining effort for both data harmonization and data integration (recovering missing information on institutions and publications). Hence, we made first a thorough effort to harmonize institution names across individual records and across data editions, by assigning them a unique identifier. Second, we conducted an online search in order to geo-localize as many institutions as possible. This involved some meticulous side-work combining web scraping and manual checking efforts. Third, we harmonized RSA researchers' identifiers in both data editions – albeit only for the sample used in this paper – by assigning a single identifier to the researchers who originally had two. Fourth, we proceeded with identifying researchers with a migratory background, based on their education and professional history.

We then moved to recover the references for such publications, based on the Web of Science (WoS) database. The WoS is one of the largest citation database, originally produced by the Institute for Scientific Information (ISI) and now maintained by Clarivate Analytics. It provides a comprehensive coverage of the scientific literature, by linking multiple databases across a wide range of research areas, for more than 33,000 journals. Its coverage spans to over 100 years of fully indexed contents, including over 1 billion records and back files (Clarivate Analytics, 2017).

In order to retrieve the reference lists of our interest, we first conducted a search based on the researchers' names and publication titles, as found in the NRF dataset. However, we got limited results, since for a simple query of this type to be successful one should have the exact name of authors and publication titles, which is far from being the case in the NRF raw data.

Therefore, we adopted a different approach. We first engaged into a long and tedious work of cleaning and standardizing the publication titles and authors' names in the 214,242 publications in our NRF-based sample. In parallel, we downloaded from the WoS all publications records from 1975 to 2014 that reported at least one author with a RSA affiliation, for a total of over 200,000 records. Last, we matched the publications in our NRF-based sample to those we downloaded for the WoS, based upon fuzzy matching of authors' surnames, name initials, title, and year of the publications. In this way, we were able to match almost 90,000 items, that is 42% publications in the NRF-based sample – for 3,713 RSA researchers out of the over 4,000 researchers of the initial NRF data. All these publications report at least one author's affiliation located in RSA, and have been produced in RSA either

by migrant researchers – before migration or after return – or by non-migrant ones. For the sake of simplicity, we refer to these data as the “RSA-based publications”.

As for the remaining publications from the NRF database that we could not match with WoS, (around 124,000), we assumed them to have been written by RSA researchers while abroad, and therefore not to be considered as part of returnee scientists’ production upon return, nor produced by non-migrants. A manual control on a small sample of these unmatched publications confirmed they are mostly articles published by RSA researchers under foreign affiliations⁴⁵. Since our goal for the present paper is to assess how the returnee researchers exploit their migration experience in their home country, we decided to set these publications aside⁴⁶.

For the present analysis, we then focus on the RSA-based publication data. For each publication, we downloaded from WoS the entire reference list, and extracted their details, as necessary for our analysis.

4.3.2.2 Migration pattern

The NRF data concern all RSA researchers applying to be rated for funding, irrespective of their nationality for this application, as long as they are affiliated to a RSA institution. The vast majority of them are RSA citizens – 2,953 out of the 3,713 researchers in our matched sample, so around 80%. Therefore in the present analysis, when speaking of “RSA researchers” we will refer exclusively to the RSA nationals in our sample, and exclude all other researchers⁴⁷.

⁴⁵ Within this group, there are researchers who have never published under a RSA affiliation. These are mainly researchers with a foreign citizenship. They are around 756 researchers in the NRF data, with over 20,000 publication references. We can easily omit these researchers as including them would bring in some bias into our analysis, since we might not be able to disentangle the real effect of migration from any effect related to some potential network that a foreign citizenship could give a researcher access to.

⁴⁶We acknowledge the limitation of using this reduced sample of publication references since we are omitting an important part of research output from RSA researchers; which includes publications made during their experience abroad or under a foreign affiliation. However, we believe this omission wouldn’t bias our analysis or at least its findings insofar that the present discussion is centred on the impact of migration on the RSA research environment. We are currently working on building a strategy to retrieve these unmatched publications or publications from RSA researchers flagged with a foreign affiliation. Besides, there are a few other publications – slightly less than 5% - which are simply not recorded in WoS. For the latter group, it could be a cumbersome task – and even not possible – to recover full relevant publication details.

⁴⁷ Direct residency permit can be granted to people who have been working or living in RSA for a minimum period of 5 years, to the spouse or dependent of RSA citizens/permanent residence permit holders. Besides, there is residency-on-other-ground permit which can be granted to people with a permanent work offer in RSA or with

Due to our focus on return migrants, we are particularly interested in the duration of their migration spell, whose measurement hides several conceptual and methodological traps, widely discussed in the migration literature – we propose a discussion on the conceptual definition of return migrants in Appendix F. A glance at our data shows around 80% of migration spells in our sample lasted from 1 to 5 years⁴⁸, with a limited number of cases where that period extended from 10 to 19 years. This suggests we are in presence of relatively short-term mobility phenomena, possibly related to postgraduate education, post-doc working experiences or visiting stays in foreign host institutions.

It should be noted that a certain number of researchers in our sample, albeit applying for funding in RSA, appear to have been abroad during their entire professional history. For this reason we exclude them from our sample, which further reduces it to 2,807 for almost 50 thousands publications. For data quality reasons, we decide to restrict our analysis to the period 1985-2014.

In Table 4.1, we report figures on the international mobility of researchers in our sample, between 1985 and 2014. We distinguish between work and education experience, with the latter limited to getting a Master degree abroad. Notice that we consider getting a PhD abroad as part of the work experience since two third of the people who did their PhD abroad had already acquired some work experience abroad with some publication before starting their PhD – we present a discussion about introducing a post-PhD migration dummy and the potential bias it might bring into our analysis in Appendix F. However for comparison reasons, we also report figures when distinguishing between PhD and other work experiences abroad and in the RSA in Table 4.1.

There is a total of 1,145 RSA researchers with past international mobility for work purpose – around 41% of the total RSA researchers of our sample (350 of which hold a foreign Master degree). For the sake of our analysis we consider those to be returnee migrants. Instead, 1,662 RSA researchers do not have any record of past work experience abroad– only 130 of them have obtained their Master degree abroad –, who constitute our non-migrant group.

exceptional skills or intend to establish a business in RSA or qualify as a refugee or qualify as a retired person or are financially independent or is a relative of a RSA citizen/ permanent residence permit holders. There are 172 individuals with permanent residence among which 106 individuals did their undergraduate in RSA and only 66 abroad. We treat these individuals as RSA researchers.

⁴⁸In a few cases a single appointment/the sum of appointments lasted for less than a year. However, this only applies to 31 researchers in the total 2,807 researchers of our sample. There is no major change in our results when we do not include them in our regressions as shown in Table F1 from Appendix F. Therefore, we have enough confidence to keep them and to define the length of stay as appointments that lasted for at least one year abroad.

Table 4.1: RSA returnees 1985-2014: PhD, Master degree and work experience

Master degree:				
Work experience:		Abroad	In RSA	Total
Abroad		350	795	1,145
In RSA only		130	1,532	1,662
Total		480	2,327	2,807
PhD degree:				
Work experience:	No PhD	Abroad	In RSA	Total
Abroad	66	356	723	1,145
In RSA only	167	166	1,329	1,662
Total	233	522	2,052	2,807

4.3.2.3 Returnees' vs. non-migrants' characteristics

Table 4.2 below shows some key characteristics of the researchers in our sample. The sampled RSA researchers are predominantly male and over 78% of them are white. We can also see that the percentage of male returnees over the total male population is more important than the same indicator for females – 44% for males against 33% for females. The mean average year of birth of researchers is higher for non-migrants as compared with returnees, but not much; we get 1955 for returnees and 1958 for non-migrants.

A sizeable share of RSA returnees is oriented towards Life sciences than in any other of the 6 research areas; about 41% for a total of 1,159 individuals. This research area also represents the area with the highest intake of returnees and non-migrants: 492 and 667 respectively.

Table 4.2: Sample characteristics

Variable	Total (N=2,807)	Returnees (N=1,145)	Non-migrants (N=1,662)
Year of birth (mean)	1957	1955	1958
Gender			
Male	1,917	849	1,068
Female	890	296	594
Race			
White	2,196	850	1,346
Others	611	295	316
Field of research			
Humanity and Arts	337	112	225
Social sciences	362	124	238
Technology	315	143	172
Medical sciences	373	142	231
Physical sciences	261	132	129
Life sciences	1159	492	667

The distribution of destination countries is highly skewed, with a few countries accounting for a very large share of migrants. The first three columns of Table 4.3 below present the top 15 countries where RSA returnees have spent the most of their work experience abroad between 1985–2014. The first two lines show around 35% and 29% of RSA returnees have worked in the US and the UK respectively in the past. There are only four African countries in this list, with Zimbabwe being the highest ranked, but only 5% of RSA returnees having a past work experience there. When looking at the list of the top 15 host countries for education – those where RSA researchers got their Master degree – in the last three columns of Table 4.3, we see here again the US and the UK appearing at the top. Both countries account for 19% of the total Master degree earned abroad. However, the distribution seems to be more even. Besides, a few more African countries get a better ranking in this list. For instance, Nigeria which is positioned third, for 5% of Master earned abroad.

Table 4.3: Top host countries between 1985-2014

Work experience abroad			Master degree obtained abroad		
Country	Nr. Researchers	Freq. (%)	Country	Nr. Researchers	Freq. (%)
USA	402	35	USA	52	11
UK	335	29	UK	38	8
Germany	113	10	Nigeria	25	5
Australia	70	6	Zimbabwe	13	3
Zimbabwe	58	5	Germany	10	2
Canada	52	4	Kenya	9	2
Netherlands	47	4	Russian Feder.	8	2
France	40	3	Poland	7	1
Nigeria	37	3	India	6	1
Namibia	30	3	Australia	5	1
Sweden	28	2	Namibia	5	1
Kenya	25	2	Malawi	4	1
Belgium	24	2	Switzerland	4	1
Switzerland	24	2	Belgium	3	0.5
Italy	18	1	Canada	3	0.5

4.4 Results

4.4.1 Baseline results: panel data analysis

The dependent variable in our regressions is the recency of publications cited by RSA researchers in the peer-reviewed journal articles the published while active in RSA. This is meant to capture how close these researchers' outputs are to the latest development/findings in the literature or in their field of research⁴⁹. We operationalize this variable with two alternative indicators.

The first indicator represents the average age of each RSA researcher's citations relative to the observed year. It is computed as the average per year and per RSA researchers of the difference between an article publication year and the average year of its backward citations. Equation (4.2) below presents the formula we use to compute this variable:

⁴⁹ In the literature, preferentially citing recent prior arts – along with citing seminal work from the past – has been depicted as a signal to the future impact of new knowledge (Mukherjee, Romero, Jones, & Uzzi, 2017), thus its position in the knowledge frontier

$$R_{it} = t - \frac{\sum \sum y_{rj} / n_j}{N_{jit}} \quad (4.2)$$

where R_{it} is the average age of citations from researcher i in year t ; t is the publication year of publication j ; y_{rj} is the publication year of reference r cited by publication j ; n_j denotes the total number of cited references per publication j ; and N_{jit} is the total number of publications j per year t , per individual i . The closer R_{it} is to 0, the more recent – or younger – are references cited by a RSA researcher in his/her publications in a given year. Table 4.4 below illustrates how this first recency indicator was computed for two RSA researchers with publications in 1990, 1993, 1994 and 2005. The last column of the table shows the values of the average age of citations of these two researchers for each year.

Table 4.4: Sample publications with the average age of citations.

Researcher ID	Publication ID	Publication Year	Avg. citation age, per publication	Avg. citation age, per year
2	WOS:A1990DU28300006	1990	11.45	10.19
2	WOS:A1990DK71200008	1990	8.92	
2	WOS:000227503100046	2005	7.72	7.9
2	WOS:000230707800012	2005	8.08	
5	WOS:A1993KY96600011	1993	14.07	9.5
5	WOS:A1993PD64600008	1993	4.67	
5	WOS:A1993LW59500002	1993	7.27	
5	WOS:A1993LL68000033	1993	12	
5	WOS:A1994QE74600004	1994	5.5	
5	WOS:A1994NZ14000006	1994	8.5	10.14
5	WOS:A1994PV94200002	1994	19.20	
5	WOS:A1994PK85900004	1994	7.37	

In Figure 4.1 we plot the yearly average age of citations for all RSA researchers. We can see the average age has been increasing over the years as a general trend. This may be an indication of the NRF data variability in its research field coverage over the years since the NRF rating started with Natural Sciences to slowly open up to other scientific fields over the years – to account for this variability in our recency indicators, we normalize them by research field average and run the baseline regressions whose results are shown in Appendix G. This increasing trend could also be due to the easiness in accessing older publications resulting from the digitalization of resources. All this suggests it is important to control for

publications year in order to account for any underlying factor likely to affect the yearly trend of citations age that we observe in this figure.

Figure 4.1: Trend in the average age of citations 1985-2014

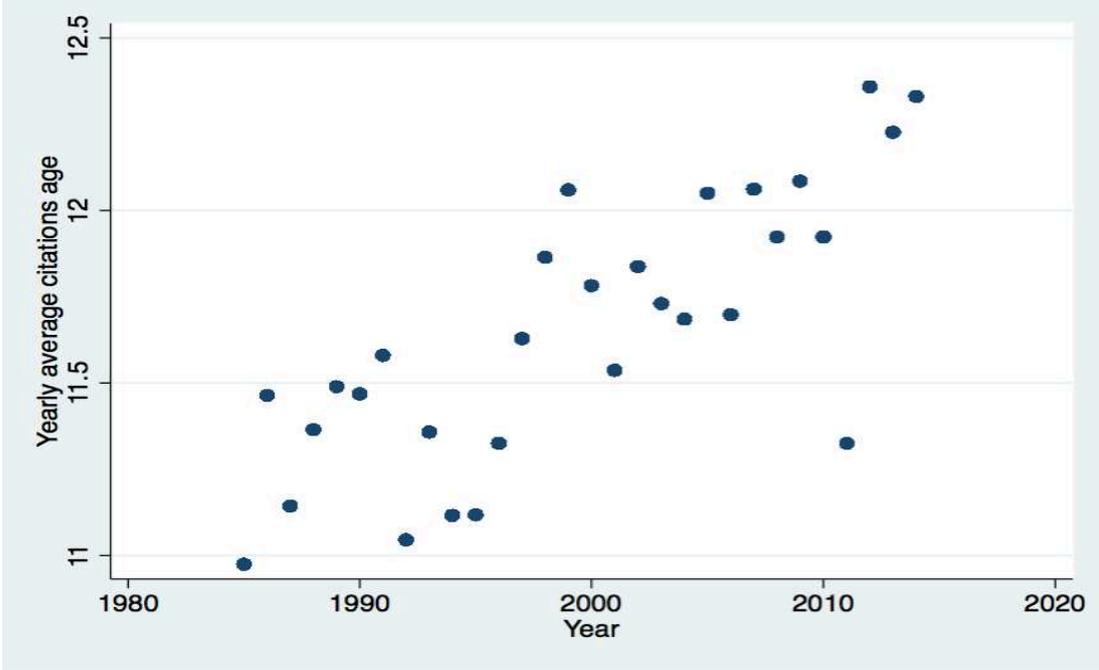
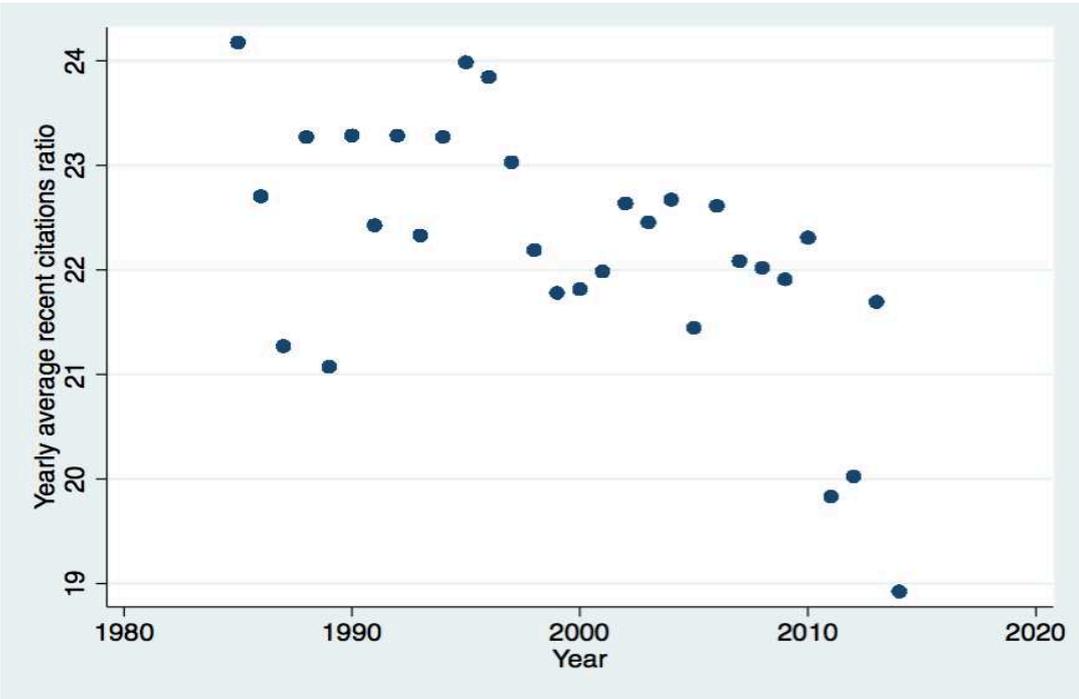


Figure 4.2: Trend in the average recent citations ratio in % 1985-2014



Our second recency indicator is the share of citations whose publication years fall within a 3-years time frame of the current publication year, averaged by year – which we will call recent citations ratio henceforth. To compute this variable, we count the number of backward citations in a given article, which are less than three years older than the citing one; then we divide this number by the total number of citations in the article; finally, we calculate the average of this ratio across all the publications by a single RSA researcher, in the observed year.

Figure 4.2 depicts the trend of the mean of the average recent citations ratio in percentages for all RSA researchers. A glance at the graph shows there is a declining tendency of this mean ratio from 1985 to 2014. Similarly to the average citations age trend, this trend could be reflecting either an imbalanced coverage of the NFR data throughout the years, or just an easiness of access to scientific resources brought up by the digitalization of these resources.

Our main explanatory variable is a dummy taking value 1 for the years following any period the RSA researcher has recorded as time spent abroad working for a non-RSA institution in his/her CV, and value 0 otherwise. All the years corresponding to the time spent abroad are dropped from our data – even if the RSA researcher had recorded to be on a double affiliation in RSA and abroad during the same period – so that to be certain to solely capture the returnee effect in a RSA research environment. This variable represents past migration experience for work purpose as opposed to past migration experience for education. Therefore, we build another variable to capture the latter effect. That is, we compute a variable taking value 1 for the years following the years of Master degree obtained abroad and 0 otherwise.

We control for individual time-invariant and time-varying characteristics of the researchers. All these variables come from information recorded in the NFR and WoS data.

The time-invariant characteristics are the researchers' year of birth, gender, race and research field. As for the time-varying ones they are:

- The accumulated number of publications per researcher, up to the observed year. This variable controls for past productivity, under the assumption that more productive researchers would tend to be more aware of up to date literature in their field of knowledge.
- The number of publications within four years before the year of PhD graduation, which controls for a possible self-selection effect into migration. Researchers who

start publishing substantially earlier in their career may produce research output that is closer to the latest development in their field of research, which in turn help them migrating.

- The number of years with a recorded job appointment up to the observed year. This variable is meant to capture the researchers' work experience or professional seniority. The expected sign of the coefficient for this variable is not very intuitive. On one hand, one could argue that senior researchers have access to more resources, including new knowledge; thus seniority might be positively correlated with recency. On the other hand, more senior researchers may rely on an obsolete stock of knowledge accumulated throughout their entire career and hence tend to cite less recent literature.

Table 4.5: Variables definition

Variables	Definition
Dependent variables	
Average age of citations	Variable measuring citations age of researcher i in year t . It is the average per year and per RSA researchers of the difference between an article publication year and the average year of its backward citation references (age). The closer this variable value is to 0, the more recent are citations of an individual during the observed year t .
Ratio of citations (\geq year -3) in %	Variable measuring the percentage in the total citations of the most recent citations of researcher i in year t . It is computed as the ratio of citations whose publication years fall within a 3-years time frame of the current publication year t , averaged by year for each researcher i .
Explanatory and control variables	
Past work migration dummy	Dummy taking value 1 for the years following an individual i working period out of RSA and 0 otherwise.
Past Master abroad dummy	Dummy taking value 1 for the years following an individual i year of Master degree obtained abroad and 0 otherwise.
Year of birth	Dummies for each decade corresponding to researcher's i birth year. There are 7 of these dummies starting from the 80s' up to the 20s' and after.
Female	Dummy taking value 1 when individual i , is a female and 0 otherwise.
White	Dummy taking value 1 when individual i is white and 0 otherwise,
Field of research	Dummies for each research area. There are 6 of them. A dummy takes value 1 individual i is specialised in that research area and 0 otherwise.
Past productivity	Accumulated number of publications made before the observed year t by researcher i .
Early publications	Number of publications within four years before the year of PhD obtained
Years of experience	Number years of work experience before the observed year t .

Table 4.6 below reports the results from the random effect model explaining the recency of RSA researchers' citations. It shows outcomes from the regressions on each of our two

dependent variables; average citations age in columns (1) to (3) and recent citations ratio in columns (4) to (6).

Outcomes from the first baseline specification, with *citations' average age* dependent variable, are shown in column (1). The first estimator captures the impact of our variable of interest – post work migration dummy – on the average citations age. It indicates that the average age of citations relative to the observed publication year decreases by 0.454 year for RSA returnees, compared to both their pre-migration publications and the publications of non-migrants.

One interesting result is the estimator for the variable *female* which is negative and statistically significant with a value of -1.067. This means women are more likely to cite more recent literature than men. We also find a negative and significant coefficient for the variable *early publication* representing the number of publications within four years before the year of PhD, which has a value of -0.026. This result implies the earlier in his/her career a researcher starts publishing a substantial number of papers, the highest his/her propensity to cite recent references in his/her future publications.

Besides, we find a significant value of 0.048 for the estimated coefficient of the variable *years of experience*. This goes against the assumption that senior researchers will tend to build their research upon more recent knowledge, due to their easier access to new materials. In fact, senior researchers appear to be more likely to cite older literature. One explanation of this result is that senior researchers have a higher incentive to build their research upon their past research. Therefore, they will likely cite their past work or the output they have produced throughout their entire career, beside the literature they have been exposed to over the years. This might not be the case for younger researchers who are still at the start of their career and wish to build their work on contemporary or trending topics.

In column (2), we add to the baseline model the dummy variable for education-related migration *master abroad*. We find no impact of this variable on the average citations age. One major implication of this result is that undergraduate studies abroad, being far from research, do not matter for recency as captured by citations average age. There are other types of experiences that are more important; this is mainly a combination of PhD, post-doc and other related work experiences proxied by our post work migration dummy. As for the other estimators, we observe no major change in their value and significance.

Table 4.6: Return migration and citations recency

Column	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Avg. age of citations			Recent citations ratio in %		
Post migration (<i>Mig_{it}</i>)	-0.454*** (0.150)	-0.446*** (0.155)		1.824*** (0.436)	1.728*** (0.454)	
Master abroad		-0.049 (0.245)			0.545 (0.695)	
Past migration in the US			-0.694*** (0.189)			1.713*** (0.612)
Past migration in the UK			-0.595** (0.267)			2.573*** (0.697)
Past migration in Africa			-0.033 (0.378)			1.141 (0.957)
Past migration to ROW			-0.073 (0.220)			0.348 (0.642)
Female	-1.067*** (0.191)	-1.067*** (0.191)	-1.054*** (0.191)	0.554 (0.530)	0.559 (0.530)	0.506 (0.530)
White	0.137 (0.210)	0.129 (0.211)	0.172 (0.214)	1.620*** (0.601)	1.710*** (0.610)	1.613*** (0.620)
Humanity and Arts	1.477*** (0.433)	1.482*** (0.435)	1.471*** (0.431)	0.806 (0.968)	0.752 (0.966)	0.770 (0.969)
Social sciences	-2.873*** (0.272)	-2.868*** (0.273)	-2.866*** (0.272)	6.264*** (0.867)	6.208*** (0.867)	6.236*** (0.866)
Technology	-2.372*** (0.275)	-2.368*** (0.275)	-2.393*** (0.275)	6.762*** (0.881)	6.722*** (0.884)	6.849*** (0.881)
Medical sciences	-2.746*** (0.224)	-2.743*** (0.224)	-2.737*** (0.224)	5.501*** (0.698)	5.467*** (0.700)	5.480*** (0.698)
Physical sciences	0.324 (0.313)	0.324 (0.313)	0.319 (0.313)	1.683** (0.794)	1.678** (0.794)	1.748** (0.796)
Life sciences	--	--	--	--	--	--
Past productivity	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	0.005 (0.006)	0.005 (0.006)	0.005 (0.006)
Early publication	-0.026** (0.012)	-0.026** (0.012)	-0.027** (0.012)	0.086** (0.040)	0.086** (0.040)	0.090** (0.040)
Years of experience	0.048*** (0.018)	0.048*** (0.018)	0.049*** (0.018)	-0.112** (0.047)	-0.110** (0.048)	-0.118** (0.048)
Constant	10.68*** (0.529)	10.69*** (0.531)	10.63*** (0.532)	20.98*** (1.678)	20.85*** (1.681)	21.71***(1.685)
Observations	19,209	19,209	19,209	19,209	19,209	19,209
R-squared within	0.0257	0.0258	0.0258	0.0053	0.0053	0.0052
R-squared between	0.1150	0.1150	0.1170	0.0471	0.0474	0.0492
R-squared overall	0.0804	0.0804	0.0819	0.0356	0.0358	0.0366
No. of researchers	2,807	2,807	2,807	2,807	2,807	2,807
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth decades and dummies for researchers' scientific field.

In column (3), we split migration experiences by country of destination, with special focus on migration to the US, the UK, other African countries besides RSA and the rest of the world (ROW) respectively. These are dummy variables taking value 1 for the years following a year working abroad in the US, the UK, in an African country and the rest of the world (ROW) respectively. As with our main explanatory variable, only the years spent in RSA are considered since we dropped all the years corresponding to single affiliation abroad and double affiliation abroad and in the RSA. The results from this model specification show having a past work migration experience in the US or in the UK negatively impacts on the average citations age, with coefficient values of -0.694 and -0.595 respectively⁵⁰; while we do not find any effect of past migration to the other location on our dependent variable.

Our results from the baseline model specification with the alternative dependent variable, the recent citations ratio, are displayed in columns (4) to (6). In column (4), the estimator for the post work migration variable returns a value of 1.824 and is statistically significant. This result indicates that the percentage of less than 3-year-old citations on returnees' publications upon their return is 1.8% higher than for the same returnees before migration and non-migrant researchers. In other words, RSA returnees tend to cite 1.8% more literature whose age or publication year is closer to the current publication year.

With the new dependent variable however, we do not find any longer any impact of the female variable, while the coefficient for the variable white is significant with a positive sign (1.620). Additionally, we find the early publication variable to be positively associated with the recent citations ratio, as its statistically significant estimator returns a value of 0.086. This suggests RSA researchers who start publishing earlier in their career a substantial amount of publications tend to cite a greater percentage of the most recent papers. Besides, we find a negative relationship between the variable for the total years of experience and our dependent variable the recent citations ratio. The estimator for this variable has a value of -0.112; suggesting the ratio of the most recent cited literature is higher for younger researchers than for senior ones. Therefore, seniority only impacts negatively on the recency of citations, as suggested by the results with the previous alternative recency measurement.

Column (5) reports the results from the regression when we introduce the dummy variable for Master degree obtained abroad in the baseline model. We do not find any effect of this

⁵⁰ A t-test shows there is a statistically significant difference in the US and the UK migration dummies estimated parameters.

explanatory variable on the recent citations ratio, while other estimators remain almost unchanged.

As with the average citations age regressions, we introduce the four dummies for returnees specific locations – US, UK, Africa and ROW – and we present the results we get in column (6). The first variable has a significant estimator of 1.713, which means that RSA returnees with at least one year experience in the US tend to cite around 1.7% more literature whose publication year is closer to the current publication year. We find an estimator with a value of 2.573 for RSA returnees who have at least a one year experience in the UK⁵¹. This means these researchers will cite around 2.6% more of the most recent literature. We do not find any effect of the dummies for the years following migration experience in other African countries or in the ROW. These results are in line with the ones from column (3) for the average citations age regressions. Results for the other variables are similar to what we had in columns (4) and (5).

4.4.2 Robustness check: conditional dif-in-dif analysis

The baseline results presented in the previous section suggest a positive relationship between past migration experience and the recency of citation references. However, they are inherently descriptive, as the relationship might simply capture pure correlation between the two variables to the extent that migration exogeneity remains questionable. Indeed, the decision to migrate might be driven by unobservable individual characteristics or by unobservable institutional factors such as policy measures or funding allocation practices that escape our attention. For instance, the existence of specific grants targeting researchers with a certain level of achievement – those who would likely be at the frontier of knowledge – may influence the decision to migrate and at the same time be available only in given years or for specific discipline, universities, or social groups.

This may create some bias in our results. Therefore, in order to overcome such problem – at least tentatively – we adopt a matching procedure where returnees (treated observations) are matched to researchers with no migration experience, but whose personal profile suggests they were potential migrants as well (control observations). In general, matching procedures of this type seek to replicate a randomized experiment in which a group of treated individuals

⁵¹ A t-test shows there is a statistically significant difference in the US and the UK migration dummies estimated parameters.

are matched on observable variables to a group of controls from which they do not differ systematically. For doing so, we use propensity score matching (PSM) approach (Stuart & Rubin, 2008) as described below, and obtain a sample of treated and control groups to which we apply conditional difference-in-difference estimation (CDID).

4.4.2.1 Propensity score matching

We first restrict our sample to RSA researchers who started their career in RSA, which is a necessary filter if we wish to match migrants to non-migrants with similar initial careers⁵². This reduces our sample of RSA returnees to 270 individuals. For each returnee, we then look for non-migrant RSA researchers with similar migration potential at the time when the returnee first left the RSA, based on a set of individual observable features. Hence we focus on non-migrants whose career started within a two-year time frame – before or after – around the returnee’s first job year. This yields a sample of 1,498 potential controls.

For these potential controls, we estimate the propensity of a researcher to belong to the treated group – i.e. to migrate and then return. The estimated model is a one-to-one nearest neighbour (PSM) without replacement and a caliper⁵³ of 0.06 corresponding to one-quarter of the standard deviation of the propensity score (Stuart & Rubin, 2008). It is based on a logistic regression of the likelihood of a researcher to migrate in a given year as a function of his/her characteristics during the period between his/her first job year and his/her first migration year – or pseudo-migration year for the controls. These characteristics are time-invariant covariates such as gender, race, year of birth, field of research and a dummy for if a researcher have had at least one non-RSA co-author during the period before migration/pseudo-migration and, as well as the average number of publications before the year of first migration – for treated individuals – and pseudo migration – for controls.

The matching logit regression results are shown in Table E3 in Appendix E. Based on them, we get a one-to-one matched sample of 223 treated and 223 controls. In order to assess whether our PSM exercise has resulted in a balance between the covariates distribution in the control and treated groups we conduct a series of tests. The first test is check the two-sample t-test for significant differences in covariates’ means (Rosenbaum and Rubin, 1985). Table

⁵²There are 434 RSA researchers who have started their career abroad out of which 173 started it with a PhD.

⁵³The caliper states the maximum allowed difference that defines a match within a range of the propensity score. Therefore, all matches not equal to or within a standard deviation of 0.06 will be dropped.

4.7 below reports the t-test for the mean values of covariates we include in the PSM regression for each of the treated and control cohorts as observed in the unmatched sample (U) and for the reduced matched sample (M). We see there is a significant difference between the treated and potential control groups (U) for all variables, and that this difference vanishes after matching (M), pointing to a more balanced distribution of these covariates in the matched treated and control groups.

Table 4.7: Descriptive statistics for the unmatched and matched samples

Variable	Values/ range	Unmatched (U) Matched (M)	Mean		%reduct		t-test	
			Treated	Control	%bias	bias	t	p>t
Year of birth	[1900 ; 1988]	U	1960	1962	-12.1		-2.12	0.034
		M	1960	1961	-5.6	53.8	-0.61	0.544
Avg. publications	[0 ; 8.55]	U	0.840	0.494	29.6		5.97	0.000
		M	0.773	0.880	-9.2	69	-1.18	0.240
Race	Integers 1-4	U	2.048	2.122	-12.6		-1.84	0.065
		M	2.048	2.040	1.3	90	0.16	0.873
Female	Integers 0-1	U	0.759	0.649	24.4		3.67	0.000
		M	0.758	0.710	10.7	56.3	1.27	0.205
Foreign co-author	Integers 0-1	U	0.915	0.688	59.3		7.91	0.000
		M	0.915	0.922	-1.9	96.7	-0.31	0.754

Additionally, we compare the size of the standardized bias between the treated and control groups before (Unmatched) and after matching (Matched). The mean and median biases are computed as summary indicators of the distribution of the bias before and after matching. Table 4.8 below shows there is a significant reduction in the mean and median standardized biases after the match.

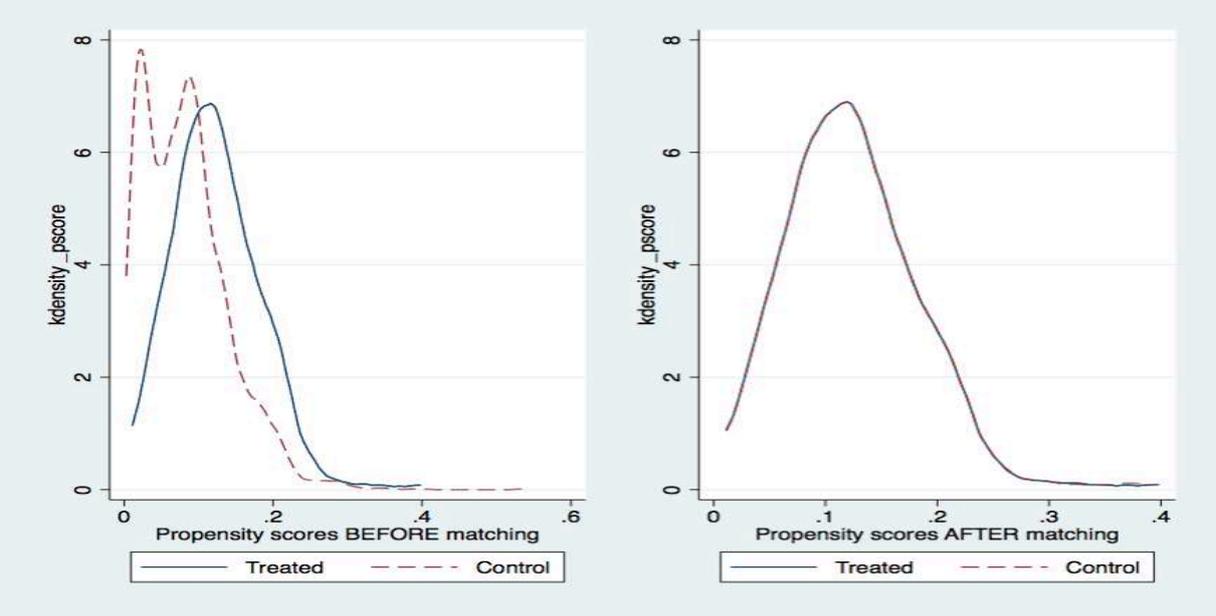
Table 4.8: Mean and Median standardized bias for the matched and unmatched samples

Sample	MeanBias	MedBias
Unmatched	26.6	23
Matched	5.4	4.6

In order to get a graphic visual of the outcome of our matching exercise, we do a comparison of the kernel density distribution of the propensity scores for the matched and unmatched samples. Figure 4.3 displays the kernel density distribution of the propensity scores before

and after the match. In the first graph depicting the propensity scores before the matching, the overlap between the graphs for the treated and control groups shows the common support condition⁵⁴ – which is one key condition for the PSM implementation – holds. On the right hand side of the figure, the curves for the treated and the control pool of researchers show these two groups are not systematically different from one another after the match. This is confirmed by Table 4.9 below which gives some detailed description of the covariates statistics after the match.

Figure 4.3: Comparison of kernel distribution of the propensity scores.



⁵⁴Also known as the overlap condition, this condition ensures that treatment observations have controlled observations ‘nearby’ or which can have a similar participation probability.

Table 4.9: Summary statistics of RSA migrant researchers and their control

	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
	Migrant researchers (223 obs.)			Matched researchers (223 obs.)		
Year of birth	1960	14.91	223	1961	13.39	223
Avg # publications	0.77	0.89	223	0.83	1.16	223
Foreign co-author	0.92	0.27	223	0.92	0.27	223
Race						
White			190			191
Black			16			17
Coloured			9			4
Indian			8			11
Gender						
Female			55			70
Male			168			153

We acknowledge the fact that our matching estimators may not be robust against some other individual unobservable characteristics – such as individual preferences or personal appreciation of risks occurring from migration – since our matching method only accounts for observable covariates. But we are confident this exercise may considerably contribute to reducing potential bias from endogeneity issue. Therefore, we use the matched sample to re-estimate the effect of treatment on citations recency.

4.4.2.2 CDID regression results

Results from the CDID regressions with year fixed effect for our reduced sample of treatment and control are presented in Table 4.10 below. Column (1) reports results from the regression on the citations average age while column (2) shows results from the regression explaining the recent citations ratio.

In column (1) we can see that the coefficient of the interaction variable $PMig_{it} * Treatment_i$, which is our CDID variable appears to be negative and significant, with a value of -0.648. We do not find any significant result for the post-migration time dummy coefficient $PMig_{it}$ and the coefficient for the treatment dummy $Treatment_i$ respectively. The result for the $PMig_{it}$ variable means we do not have enough argument supporting the assumption that there is a change in the mean average citations age from before to after pseudo-migration in the control group. The result for the $Treatment_i$ variable suggests there is no difference in the mean average citations age between the treatment and control groups prior to migration/pseudo-

migration. It is the marginal effect of belonging to the treatment group before migration, which is not significant. To get the estimated difference in the mean average citations age between the treatment and control groups after migration, one should sum up this coefficient with the coefficient of the interaction. We can get this by computing the marginal effect after migration, whose results are returned in Table 4.11. It shows the marginal effect of belonging to the treatment group after migration is -0.643 and significant. This suggests that on average RSA returnees of our matched sample tend to cite references whose average age is 0.665 year more recent than references in non-migrants' papers. All the coefficients for the other covariates lose their significance, while the coefficient for the variable past productivity, becomes significant and with a negative sign.

Table 4.10: Results from the difference-in-difference regression

	(1)	(2)
	Avg. age of citations	Recent citations ratio in %
Post migration time ($PMig_{it}$)	-0.001 (0.281)	-2.466*** (0.856)
Treatment _i	0.005 (0.449)	-0.459 (1.278)
$PMig_{it}$ * Treatment _i	-0.648* (0.382)	3.683*** (1.233)
Female	-0.265 (0.408)	-0.535 (1.011)
White	0.527 (0.509)	-1.091 (1.488)
Humanity and Arts	2.562** (1.275)	-3.105 (2.546)
Social sciences	-2.672*** (0.639)	4.594** (2.110)
Technology	-0.763 (0.583)	4.304** (1.694)
Medical sciences	-2.361*** (0.411)	3.889*** (1.212)
Physical sciences	0.170 (0.513)	1.217 (1.310)
Life sciences	-- --	-- --
Past productivity	-0.011** (0.005)	0.027** (0.013)
Early publication	-0.026 (0.040)	0.041 (0.107)
Years of experience	0.036 (0.037)	-0.146* (0.085)
Constant	9.962*** (1.173)	29.87*** (3.964)
Observations	4,090	4,090
R-squared within	0.0379	0.0160
R-squared between	0.1320	0.0807
R-squared overall	0.0927	0.0454
Number of researchers	446	446
Year FE	Yes	Yes

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

All the regressions include dummies for the year of birth decades and dummies for researchers' scientific field.

Results displayed in column (2) for the recent citations ratio regression show the interaction term has a significant estimator with a value of 3.683. The estimator for the treatment dummy $Treatment_i$ has a value of -0.459 but is not significant. This implies we cannot find any difference in the mean of the recent citations ratio between the treatment and control groups before migration/pseudo-migration. The result for the post migration/pseudo migration time dummy $PMig_{it}$ returns a negative and significant coefficient of -2.466. This shows that in the control group, during the years following the pseudo-migration there is a declining tendency of RSA researchers to cite publications which are 3 years or less older by 2.5% less than before the pseudo-migration. The estimated difference in the mean of the recent citations ratio between the treatment and control groups after migration is displayed in Table 4.11. We can see the marginal effect of being treated after migration is 3.223 and significant. In other words, RSA returnees are likely to cite 3.2% more citations 3-years old or younger than non-migrant researchers upon their return.

As for the other covariates, we only get significant results with the variables past productivity (0.027) and years of experience (-0.146).

Table 4.11: Mean output differences in the treatment and control groups before and after migration

Column	(1)	(2)
	Avg. age of citations	Recent citations ratio in %
Before migration	0.005 (0.449)	-0.459 (1.278)
After migration	-0.643* (0.339)	3.223*** (0.896)

Overall, the CDID exercise yields estimates for the main explanatory variable with relatively higher absolute values than in the baseline model. Indeed in the latter, we found the mean difference between returnees and non-migrants post-migration to be -0.454 and 1.824 for the dependent variables average citations age and recent citations ratio respectively; while the mean difference values from the CDID regressions are -0.643 and 3.223 for the average citations age and the recent citations ratio respectively. However the average citations age is significantly weaker in the CDID regression. The different results we get from the two models may be due to the fact that in the CDID model, we use a sample made of pairs of individuals

with very close characteristics. Hence, this allows us to treat for potential bias/noise, hence increasing the precision of our estimates.

4.4.3 Migration effects over time

4.4.3.1 Delay/decay in the post-migration effect

The results we got from the previous analysis point to a positive relationship between return migration and the recency of citations. However, our migration dummy only accounts for first return event and so does not capture any other migration/return event that might have occurred after that first event. Indeed, Table 4.12 below reports the frequency distribution of migration/return events for all of our sample returnees. We can see this distribution ranges from 1 to 7. Besides, about 92% of returnees have only been abroad not more than twice. Moreover, Table 4.13 shows the mean of migration/return events is 1.375. All this implies although there might be some variability within the migration/return events distribution, most returnees do not renew their migration experience after their first or second experience.

Table 4.12: Migration/return events distribution

No. of migration/ return events	Freq.	Percent	Cum.
1	835	72.93	72.93
2	213	18.60	91.53
3	58	5.07	96.60
4	21	1.83	98.43
5	13	1.14	99.57
6	3	0.26	99.83
7	2	0.17	100.00
Total	1,145	100.00	

Table 4.13: Summary statistics of migration/return

Variable	Obs.	Mean	Std. Dev.	Min.	Max
Migration/return	1,145	1.375	0.7719	1	7

Yet, omitting this variability in the migration/return frequency may cause some bias in our analysis that might occur for two main reasons. First, higher subsequent migration/return

frequencies might hide underlying factors likely to lead our positive results. And secondly in cases where we have lower frequencies of migration/return – 1 to 2 events –, our results suggest the recency effect is constant throughout the entire period from a researchers' last return year to the last observed year 2014; which might not be entirely true. One way to assess this is to allow a decay in the return migration effect for different time windows after the return year. For doing so, we use two different approaches to build new post migration dummies.

The first approach lies on building dummies taking value 1 for only the first three to six⁵⁵ years post-migration respectively – and not for all the post-migration years as we did in the main analysis. Then we run new regressions using these variables separately in the baseline model in equation (4.1). The results from these regressions for each of the two recency indicators are presented in Table 4.14 below. A glance at the post migration dummy estimators in all columns shows the closest we get to the return year, the less we find any effect of migration experience on recency. That is, that effect totally disappears within three years post migration. This result is surprising to the extent that one would expect the opposite. Indeed, if we follow the assumption that returnees recency effect is driven by their past exposure to up to date knowledge during their migration experience, we would expect them to have a higher propensity to cite more recent literature during the first years following their return and that this effect would decay the farther they get from that exposure. However, one needs to point out to the fact that returnees productivity is lower during the first years post return as they might need some time to adjust to their new environment in the RSA and start publishing some substantial amount of papers⁵⁶. Hence as productivity is positively associated with citations recency, this lack of effect during the first three years post return may in fact reflect the returnees low productivity during that period.

Overall, results from Table 4.14 suggest there might be some 3-year delay in the positive effect of post migration on recency. However, the question remains on whether this positive effect is constant or lasts throughout the returnees' post migration period after these 3 years delay. To investigate this, we use a second approach which consists on building post-migration dummies differently than with the above first approach, namely we restrict the post-migration dummies to a 3-year time window and allow that time window to move along

⁵⁵Six represents the average year gap between the first and second migration/return event, for the returnees who have been abroad more than once.

⁵⁶The average number of papers produced within the first three years upon return is around 2, for around 3.6 papers for the entire return period.

returnees' post-migration period, starting from the second year. More precisely, we build separate dummies; the first one takes value 1 for the second to fourth year post migration, then the second one is assigned value 1 for the third to fifth year post-migration, and so on. Then, we run separate regressions using these new dummies in the baseline model. Results from these regressions are shown in Table 4.15 below. In general, these results show the effect of migration experience on the average citations age start disappearing from the seventh year post-migration; while this effect on the recent citations ratio start vanishing from the ninth year post-migration. Therefore, we can assume there is a recency decay effect after a certain number of years post-migration; which might be explained by the fact that most of these returnees do not renew their migration experience after the first experience. Hence, they are likely to lose their foreign connection and network, with their knowledge exchange scope restricted to their RSA counterparts. This might in turn limit their sight on new knowledge in their field of research.

Table 4.14: Results from regressions with post migration effect delay

Colonne1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1-3 years		1-4 years		1-5 years		1-6 years	
VARIABLES	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %
Post migration (Mig_{it})	-0.177 (0.108)	0.473 (0.375)	-0.230** (0.105)	0.633* (0.357)	-0.348*** (0.107)	0.966*** (0.347)	-0.391*** (0.106)	1.361*** (0.344)
Female	-1.037*** (0.190)	0.428 (0.528)	-1.040*** (0.190)	0.438 (0.528)	-1.048*** (0.191)	0.459 (0.528)	-1.053*** (0.191)	0.489 (0.529)
White	0.161 (0.210)	1.489** (0.601)	0.154 (0.210)	1.511** (0.601)	0.139 (0.210)	1.552*** (0.601)	0.135 (0.210)	1.597*** (0.600)
Humanity and Arts	1.530*** (0.433)	0.619 (0.970)	1.528*** (0.432)	0.623 (0.969)	1.524*** (0.432)	0.633 (0.969)	1.515*** (0.432)	0.657 (0.969)
Social sciences	-2.831*** (0.272)	6.114*** (0.865)	-2.832*** (0.272)	6.117*** (0.865)	-2.836*** (0.272)	6.126*** (0.865)	-2.843*** (0.272)	6.146*** (0.865)
Technology	-2.372*** (0.275)	6.774*** (0.880)	-2.370*** (0.275)	6.767*** (0.880)	-2.364*** (0.275)	6.751*** (0.880)	-2.362*** (0.275)	6.733*** (0.880)
Medical sciences	-2.723*** (0.224)	5.397*** (0.698)	-2.726*** (0.224)	5.406*** (0.697)	-2.733*** (0.224)	5.424*** (0.697)	-2.735*** (0.224)	5.445*** (0.697)
Physical sciences	0.293 (0.312)	1.812** (0.794)	0.294 (0.312)	1.808** (0.794)	0.299 (0.312)	1.795** (0.794)	0.305 (0.312)	1.768** (0.794)
Life sciences	-	-	-	-	-	-	-	-
Past productivity	0.001 (0.002)	0.006 (0.006)	0.001 (0.002)	0.007 (0.006)	-0.001 (0.002)	0.008 (0.006)	-0.001 (0.002)	0.010 (0.006)
Early publication	-0.0273** (0.0117)	0.0875** (0.0395)	-0.0274** (0.0117)	0.0878** (0.0395)	-0.0276** (0.0117)	0.0885** (0.0395)	-0.0270** (0.0117)	0.0876** (0.0395)
Years of experience	0.0480*** (0.0181)	-0.111** (0.0475)	0.0476*** (0.0181)	-0.110** (0.0475)	0.047*** (0.0181)	-0.109** (0.0475)	0.0466** (0.0181)	-0.107** (0.0475)

Constant	10.59*** (0.526)	21.49*** (1.678)	10.62*** (0.526)	21.40*** (1.679)	10.70*** (0.526)	21.20*** (1.680)	10.72*** (0.527)	20.97*** (1.681)
Observations	19,209	19,209	19,209	19,209	19,209	19,209	19,209	19,209
No. of researchers	2,807	2,807	2,807	2,807	2,807	2,807	2,807	2,807
R-squared within	0.0253	0.00480	0.0254	0.00479	0.0258	0.00489	0.0259	0.00517
R-squared overall	0.0788	0.0333	0.0791	0.0337	0.0798	0.0345	0.0801	0.0355
R-squared between	0.116	0.0452	0.116	0.0456	0.116	0.0464	0.116	0.0472

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth decades and dummies for researchers' scientific field.

Table 4.15: Results from regressions with post migration effect decay

Colonne1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	year 2-4 post-migration		year 3-5 post-migration		year 4-6 post-migration		year 5-7 post-migration		year 6-8 post-migration	
VARIABLES	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %	Avg. age of citations	Recent citations ratio in %
Post migration (Mig_{it})	-0.200*	0.816**	-0.274**	0.831**	-0.296***	1.076***	-0.147	0.750**	-0.110	0.776*
	(0.104)	(0.358)	(0.107)	(0.359)	(0.110)	(0.372)	(0.116)	(0.381)	(0.122)	(0.399)
Female	-1.038***	0.439	-1.039***	0.439	-1.040***	0.445	-1.036***	0.434	-1.034***	0.430
	(0.190)	(0.529)	(0.190)	(0.529)	(0.190)	(0.529)	(0.190)	(0.529)	(0.190)	(0.529)
White	0.166	1.499**	0.169	1.475**	0.174	1.466**	0.179	1.451**	0.180	1.448**
	(0.210)	(0.600)	(0.210)	(0.600)	(0.210)	(0.600)	(0.210)	(0.600)	(0.210)	(0.600)
Humanity and Arts	1.523***	0.632	1.513***	0.666	1.509***	0.684	1.514***	0.672	1.516***	0.667
	(0.432)	(0.969)	(0.432)	(0.969)	(0.432)	(0.969)	(0.432)	(0.969)	(0.432)	(0.969)
Social sciences	-2.836***	6.122***	-2.841***	6.140***	-2.844***	6.154***	-2.841***	6.147***	-2.840***	6.143***
	(0.272)	(0.865)	(0.272)	(0.866)	(0.272)	(0.866)	(0.272)	(0.866)	(0.272)	(0.867)
Technology	-2.372***	6.766***	-2.371***	6.768***	-2.372***	6.767***	-2.376***	6.779***	-2.378***	6.782***
	(0.275)	(0.881)	(0.275)	(0.881)	(0.275)	(0.882)	(0.274)	(0.881)	(0.274)	(0.882)
Medical sciences	-2.722***	5.404***	-2.722***	5.397***	-2.720***	5.395***	-2.717***	5.389***	-2.717***	5.390***
	(0.224)	(0.698)	(0.224)	(0.698)	(0.224)	(0.699)	(0.224)	(0.699)	(0.224)	(0.699)
Physical sciences	0.295	1.803**	0.299	1.795**	0.302	1.779**	0.298	1.786**	0.297	1.786**
	(0.312)	(0.794)	(0.312)	(0.795)	(0.312)	(0.795)	(0.312)	(0.795)	(0.312)	(0.794)
Life sciences	-	-	-	-	-	-	-	-	-	-
Past productivity	4.15e-05	0.00707	-0.000147	0.00713	-0.000167	0.00746	0.000174	0.00663	0.000280	0.00639
	(0.00190)	(0.00582)	(0.00190)	(0.00581)	(0.00188)	(0.00583)	(0.00188)	(0.00583)	(0.00188)	(0.00583)
Early publication	-0.0270**	0.0869**	-0.0267**	0.0858**	-0.0262**	0.0843**	-0.0263**	0.0840**	-0.0264**	0.0833**
	(0.0117)	(0.0395)	(0.0117)	(0.0394)	(0.0117)	(0.0394)	(0.0117)	(0.0394)	(0.0117)	(0.0395)
Years of experience	0.0482***	-0.111**	0.0484***	-0.112**	0.0484***	-0.112**	0.0488***	-0.113**	0.0489***	-0.114**

	(0.0181)	(0.0475)	(0.0181)	(0.0475)	(0.0181)	(0.0475)	(0.0181)	(0.0475)	(0.0181)	(0.0475)
	10.56***	21.48***	10.55***	21.57***	10.53***	21.62***	10.51***	21.68***	10.51***	21.70***
Constant	(0.524)	(1.668)	(0.525)	(1.667)	(0.526)	(1.670)	(0.525)	(1.667)	(0.525)	(1.666)
Observations	19,209	19,209	19,209	19,209	19,209	19,209	19,209	19,209	19,209	19,209
No. of researchers	2,807	2,807	2,807	2,807	2,807	2,807	2,807	2,807	2,807	2,807
R-squared within	0.0254	0.00498	0.0256	0.00505	0.0256	0.00523	0.0253	0.00501	0.0253	0.00503
R-squared overall	0.0788	0.0337	0.0789	0.0336	0.0789	0.0338	0.0784	0.0333	0.0784	0.0333
R-squared between	0.116	0.0453	0.116	0.0449	0.116	0.0448	0.116	0.0446	0.116	0.0445

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth decades and dummies for researchers' scientific field.

4.4.3.2 First and subsequent migration events differentiated

In order to further assess migration effects on recency over time, we check for whether the migration experience has a long lasting effect or not. That is, we want to test if one migration event suffices to impact on recency over time or further migration experience has an independent effect. There are 310 returnees with at least two migration/return events. However, the duration of stay is not equally distributed across these migration/return events. Indeed, Table 4.16 below displays this distribution to be skewed, with most of the migration spell spanning between 1 to 5 years either for returnees with only one or more than one migration experiences. The duration of stay seems to be representative of the type of experience acquired by the returnee abroad. In general, we can see from Table 4.16 that the first migration/return experience lasts longer than subsequent migration/return events.

Table 4.16: Distribution of the number of migration/return events by the duration of stay abroad

	No. of migration/ return events		
	1	2+	Total
Duration of stay abroad			
1	239	407	646
2	79	74	153
3	54	64	118
4	238	101	339
5	43	28	71
6	34	14	48
7	27	13	40
8	16	11	27
9	11	7	18
10	8	6	14
11	14	6	20
12	11	5	16
13	16	1	17
14	10	6	16
15	11	8	19
16	8	5	13
17	8	3	11
18	6	4	10
19	4	3	7
Total	835	766	1,601

Therefore in order to assess whether these differences between first and subsequent migration events differently impact on recency, we build a second post-migration dummy that takes value 1 only for the years following the second trip abroad and 0 otherwise. This means the years following the first migration event have value 0 in this new variable.

We run the baseline model regression for each of the two recency indicators adding this new post-migration dummy and present results in Table 4.17. We do not find any effect of this new variable on recency. This may imply the first migration episode is all that matters, but not because it is the first one, but because it is the longer-lasting and so the one associated with the most intense experience. This may also be due to the relatively low number of returnees with more than one migration experiences – 310 as compared with 835 returnees with only one migration experience.

Table 4.17: Results from regressions with post migration dummy first and second trips.

Colonne1	(1)	(2)
VARIABLES	Avg. age of citations	Recent citations ratio in %
Post migration (after first trip)	-0.428*** (0.155)	1.697*** (0.459)
Post migration (after second trip)	-0.153 (0.248)	0.700 (0.747)
Female	-1.066*** (0.191)	0.551 (0.531)
White	0.134 (0.210)	1.631*** (0.601)
Humanity and Arts	1.474*** (0.433)	0.817 (0.969)
Social sciences	-2.877*** (0.273)	6.280*** (0.867)
Technology	-2.369*** (0.275)	6.751*** (0.881)
Medical sciences	-2.743*** (0.224)	5.487*** (0.698)
Physical sciences	0.329 (0.313)	1.662** (0.795)
Life sciences	-	-
Past productivity	0.001 (0.002)	0.005 (0.006)
Early publication	-0.0264** (0.0117)	0.0863** (0.0394)
Years of experience	0.0487*** (0.0181)	-0.114** (0.0475)
Constant	10.68*** (0.528)	21.02*** (1.678)
Observations	19,209	19,209
No. of researchers	2,807	2,807
R-squared within	0.0257	0.00527
R-squared overall	0.0808	0.0360
R-squared between	0.115	0.0474

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth decades and dummies for researchers' scientific field.

4.5 Concluding remarks

In this analysis, we have contributed to the brain drain, brain gain trade-off debate by investigating return migration and knowledge diffusion using a sample of active RSA researchers. More precisely, we assess the impact of these RSA researchers past migration experience on citations recency. We compute two main indicators, namely the average age of citations referenced in RSA researchers' publications and the ratio of citations whose years of publication falls within a three- year time window of the observed year. Our analysis relies on a rich dataset on RSA researchers from the NFR, the RSA national agency for research promotion. To fill some of the missing information from this original data we download citation references from the WoS.

Our analysis is conducted at two levels. The first level consists of running a random effect model with years fixed effect on our entire sample. At a second level, we conduct a CDID analysis with a reduced sample of treated and control group, with a prior PSM analysis to get a good fit in the matched sample of treatment and control. This second level of analysis allows us to reduce bias from potential endogeneity arising from selection issue and omitted variables.

Results from both levels suggest there is a positive effect of being a RSA returnees on citations recency, for both of our indicators. That is, we find RSA returnees tend to cite more recent literature – in the average year of publication and in the ratio of most recent publications – than non-migrants RSA researchers. These results highlight the benefits of foreign experience for researchers, in the sense that this foreign experience provides them with the opportunity to get exposed to new ideas or recent knowledge, with the capability to maintain this exposure even upon their return to their origin country. Our results also point to the non-availability of this exposure to non-migrants – even to those with very similar characteristics than returnees before they migrate. This suggests the research environment in migrants' home countries is not always open to international cooperation in a way that would allow non-migrants to integrate transnational professional networks of a similar calibre than the returnees' ones.

Besides, our results join findings from previous literature on returnees out-performing non-migrants in their scientific activities insofar that citations recency has been associated with

future publications quality. It also points to the ability of RSA returnees to perform at the knowledge frontier.

In general, this has some implications in terms of policies which should favour circular migration of researchers between home and host countries while encouraging their return to their home countries. This could be done through implementation of transnational research programmes which promote international mobility through visiting stays for instance. Besides, host countries – especially African countries – should provide researchers with adequate research environment home for an efficient exploitation of their knowledge and international network.

We are currently working on building a strategy to retrieve these unmatched publications or publications from RSA researchers flagged with a foreign affiliation, for further investigation of the brain gain effect of returnees.

Appendix E – Some tables

Table E1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Avg. age citations	19,209	11.79	5.76	0	86.53
Recent citations ratio in %	19,209	22.21	16.89	0	100
Past migration dummy	19,209	0.39	0.49	0	1
Master abroad dummy	19,209	0.14	0.35	0	1
Past migration in the US	19,209	0.15	0.36	0	1
Past migration in the UK	19,209	0.13	0.33	0	1
Past migration in Africa	19,209	0.05	0.23	0	1
Past migration to the rest of the world	19,209	0.10	0.29	0	1
Year of birth	19,209	1957	13.24	1900	1989
Female	19,209	0.26	0.44	0	1
Male	14,136	0	0	0	0
Female	5,073	1	1	1	1
White	19,209	0.84	0.36	0	1
White	16,220	1	0	1	1
Others	2,989	0	0	0	0
Field of research	19,209	--	--	--	--
Humanity and Arts	19, 209	0.05	0.23	0	1
Social sciences	19, 209	0.07	0.26	0	1
Technology	19, 209	0.11	0.31	0	1
Medical sciences	19, 209	0.16	0.37	0	1
Physical sciences	19, 209	0.12	0.33	0	1
Life sciences	19, 209	0.49	0.50	0	1
Past productivity	19, 209	17.35	24.74	1	553
Early publications	19, 209	2.24	4.12	0	86
Years of experience	19,209	19.39	10.17	0	54

Table E2: Correlation table

	3-years citations ratio	Avg. age citations	Femal e	White	Pub. numbe r	Early pub	Years of exp.	Work dumm y	Master dumm y
3-years citations ratio	1								
Avg. age citations	-0.535	1							
Female	0.018	-0.083	1						
White	0.021	0.013	0.050	1					
Pub. number	-0.001	0.042	-0.111	0.106	1				
Early pub	0.022	-0.032	-0.024	0.019	0.295	1			
Years of exp.	-0.041	0.122	-0.095	0.141	0.322	-0.138	1		
Work dummy	0.041	-0.014	-0.109	0.050	0.136	-0.028	0.133	1	
Master dummy	0.037	-0.013	-0.046	0.150	0.024	-0.016	0.081	0.276	1

Table E3: Logit results from the PSM analysis

VARIABLES	Returnee dummy
Year of birth	-0.006** (0.003)
Gender	0.184** (0.075)
Race	-0.131** (0.058)
Field of research	0.036*** (0.007)
Avg. Past productivity	0.122*** (0.033)
Foreign co-author	0.783*** (0.101)
Constant	8.492 (5.203)
Pseudo R-squared	0.074
Prob> Chi2	0.0000
Observations	3,126

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix F – Discussion on the conceptualization of returnees

One of the recommendations of the United Nations (UN) is to consider as migrants individuals with at least a one-year stay abroad. This definition is the one that is general used by many statistical offices, for census or survey purposes. Therefore if we follow it, individuals from a given country of origin who have moved to a different country of destination for work, education, or any other motive belong to the group of migrants as long as they have spent at least a year abroad. Once these migrants decide to go back to their home country, they become returnees. One definition of returnees proposed by Dustmann & Weiss (2007) is that of migrants who settle back in their home country by their own choice, after having spent several years abroad. However, this definition does not give any clear precision on the length of stay. Hence we can assume that since returnees are first and foremost migrants, the definition provided by the UN also applies to them. Furthermore, the minimum length of stay of 1 year for returnees is the one that have been adopted in most empirical paper on return migration (Jonkers & Cruz-Castro, 2013; Velema, 2012).

Admittedly, returnees might be a broad concept encompassing different types of migrations: short-term, medium term, long term. That is why a special attention needs to be paid on the length of stay in any empirical analysis on returnees. In our data, about 80% of researchers have a migration time spanning between 1 to 5 years, pointing to a rather short term migration. That is, a combination of migration for work, PhD or Post-doc. It should be noted that our data does not allow us to have the exact dates of appointment for many researchers. This means for these researchers – 31 of them –, we only have the starting year and the ending year of assignment. For the same starting and ending year, it is therefore not possible to say if those appointments covered a full year length. As a robustness check, we remove those 31 researchers from our sample and run the baseline model regressions, whose results are shown in Table F1. The results we get are not different from the results with the main sample. Hence, we have enough confident to use our whole sample and define returnees as researchers with at least a one year stay abroad.

Besides, distinguishing between migration for work purpose or for PhD does not seem to be a useful strategy insofar that most of the researchers in our data have already started their career

with some publications before taking up a PhD⁵⁷. Therefore, a post-PhD migration dummy would treat publications made by returnees from work experience abroad as publications made by these researchers before they migrate. Hence this might cause some bias in our analysis on the effect of migration experience on citations recency. However, we created a post-PhD migration dummy and ran the baseline model regressions. Results from these regressions are shown in Table F2 below.

⁵⁷There are only 173 returnees who started their career doing a PhD abroad and most of them didn't renew their migration experience after PhD.

Table F1: Baseline regressions without researchers with less than a year long migration experience

Colonne1	(1)	(2)
VARIABLES	Avg. age of citations	Recent citations ratio in %
Post migration (<i>Mig_{it}</i>)	-0.449*** (0.153)	1.882*** (0.443)
Female	-1.064*** (0.192)	0.417 (0.534)
White	0.120 (0.213)	1.739*** (0.606)
Humanity and Arts	1.486*** (0.437)	0.852 (0.978)
Social sciences	-2.878*** (0.274)	6.388*** (0.872)
Technology	-2.389*** (0.277)	6.856*** (0.889)
Medical sciences	-2.749*** (0.225)	5.668*** (0.700)
Physical sciences	0.334 (0.316)	1.911** (0.799)
Life sciences	-	-
Past productivity	-5.46e-05 (0.00187)	0.00486 (0.00585)
Early publication	-0.0257** (0.0118)	0.0747* (0.0391)
Years of experience	0.048*** (0.0182)	-0.111** (0.0478)
Constant	10.70*** (0.532)	20.97*** (1.683)
Observations	18,895	18,895
No. of researchers	2,776	2,776
R-squared within	0.0253	0.00569
R-squared overall	0.0795	0.0367
R-squared between	0.116	0.0483

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth decades and dummies for *researchers' scientific field*.

Table F2: Regressions with PhD experience abroad

Colonne	(1)	(2)
VARIABLES	Avg. age of citations	Recent citations ratio in %
PhD abroad	-0.034 (0.209)	1.443** (0.590)
Female	-1.033*** (0.190)	0.465 (0.530)
White	0.177 (0.211)	1.644*** (0.604)
Humanity and Arts	1.519*** (0.432)	0.595 (0.971)
Social sciences	-2.837*** (0.272)	6.076*** (0.866)
Technology	-2.376*** (0.276)	6.638*** (0.885)
Medical sciences	-2.717*** (0.224)	5.454*** (0.700)
Physical sciences	0.296 (0.313)	1.693** (0.795)
Life sciences	-- --	-- --
Past productivity	-0.001 (0.001)	0.005 (0.006)
Early publication	-0.027** (0.012)	0.092** (0.040)
Years of experience	0.049*** (0.018)	-0.107** (0.048)
Constant	10.51*** (0.530)	21.30*** (1.673)
Observations	19,209	19,209
No. of researchers	2,807	2,807
R-squared within	0.0253	0.0053
R-squared between	0.1150	0.0441
R-squared overall	0.0782	0.0334

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth decades and dummies for researchers' scientific field.

Appendix G – Normalizing recency over scientific field average

One specific feature of the NRF data is that there has been some variability in the research field coverage over the years⁵⁸. To account for this variability, we normalize our recency indicators over research field averages. That is, for each researcher/year unit we divide the observed recency by the average recency in the researcher's research field during the observed year. Table G1 below depicts the results from the regressions on these normalized recency indicators. Both columns show both recency indicators keep their sign and their significance.

Table G1: Results from regressions with recency normalized over scientific field average

Variable	(1) Avg. age of citations	(2) Recent citations ratio in %
Post migration (<i>Mig_{it}</i>)	-0.039*** (0.012)	0.082*** (0.020)
Female	-0.082*** (0.015)	0.025 (0.023)
White	0.003 (0.018)	0.072*** (0.027)
Past productivity	0.001 (0.002)	0.001 (0.002)
Early publication	-0.002* (0.001)	0.004** (0.002)
Years of experience	0.004** (0.002)	-0.005** (0.002)
Constant	0.851*** (0.043)	1.062*** (0.075)
Observations	19,209	19,209
No. of researchers	2,807	2,807
R-squared within	0.0243	0.00568
R-squared overall	0.0200	0.00980
	0.0164	0.00938

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All the regressions include dummies for the year of birth

decades and dummies for researchers' scientific field.

⁵⁸ NRF rating system initiated with natural sciences only, then slowly started opening to other research fields.

Bibliography

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1–19.
- Abramo, G., D'Angelo, C. A., & Solazzi, M. (2012). A bibliometric tool to assess the regional dimension of university–industry research collaborations. *Scientometrics*, 91(3), 955–975.
- Ackers, L. (2005). Promoting Scientific Mobility and Balanced Growth in the European Research Area. *Innovation*, 18, 301–317.
- Adepoju, A., van Naerssen, T., & Zoomers, A. (2008). International Migration and National Development in SubSaharan Africa: View Points and Policies Initiatives in the Countries of Origin. In Aderanti et al. (Eds.). Brill E-Books.
DOI:10.1163/ej.9789004163546.i-309.13.
- Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: Knowledge flows, labor mobility and enduring social relationships. *Journal of Economic Geography*, 6, 571–591.
- Agrawal, A., Kapur, D., & McHale, J. (2008). How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data. *Journal of Urban Economics*, 64(2), 258–269. <https://doi.org/10.1016/j.jue.2008.01.003>
- Agrawal, A., Kapur, D., McHale, J., & Oettl, A. (2011). Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation. *Journal of Urban Economics*, 69(1), 43–55.
- Almeida, P., Phene, A., & Li, S. (2010). Communities, knowledge and innovation: Indian immigrants in the US semiconductor industry. Working Paper, 58/2010 (globADVANTAGE – Center of Research in International Business & Strategy).

- Alnuaimi, T., Opsahl, T., & George, G. (2012). Innovating in the periphery: The impact of local and foreign inventor mobility on the value of Indian patents. *Research Policy*, 41(9), 1534–1543. <https://doi.org/10.1016/j.respol.2012.06.001>
- Appelt, S., van Beuzekom, B., Galindo-Rueda, F., & de Pinho, R. (2015). Which factors influence the international mobility of research scientists? In *Global Mobility of Research Scientists* (pp. 177–213).
- Arslan, C., Dumont, J.-C., Kone, Z., Moullan, Y., Ozden, C., Parsons, C., & Xenogiani, T. (2015). A new profile of migrants in the aftermath of the recent economic crisis.
- Artuç, E., Docquier, F., Özden, Ç., & Parsons, C. (2015). A global assessment of human capital mobility: the role of non-OECD destinations. The World Bank.
- Ascencio, F. L. (1993). Bringing it back home: Remittances to Mexico from migrant workers in the United States (Vol. 37). Center for Us-Mexican Studies.
- Ashenfelter, O. C., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs (Vol. 67). *Review of Economics and Statistics*.
- Auriol, L. (2007). Labour market characteristics and international mobility of doctorate holders: results for seven countries (OECD STI Working Paper). Paris: OECD Publishing.
- Auriol, L. (2010). Careers of doctorate holders: employment and mobility patterns (OECD STI Working Paper). Paris: OECD Publishing.
- Barlett, C. A., & Yoshihara, H. (1988). New Challenges for Japanese Multinationals. Is Organizational Adaptation their Achilles Hell? *Human Resource Management*, 27, 19–43.

- Barnard, H., Cowan, R., & Müller, M. (2016). On the value of foreign PhDs in the developing world: Training versus selection effects. Working Paper Series in Economics, Karlsruher Institut für Technologie (KIT).
- Barré, R., Hernandez, V., Meyer, J.-B., & Vinck, D. (2003). Scientific Diasporas. IRD.
- Barro, R. J., & Lee, J. W. (2013). A New Data Set of Educational Attainment in the World, 1950–2010. *Journal of Development Economics*, 104(C), 184–198.
- Baruffaldi, S. H., & Landoni, P. (2012). Return mobility and scientific productivity of researchers working abroad: The role of home country linkages. *Research Policy*, 41(9), 1655–1665. <https://doi.org/10.1016/j.respol.2012.04.005>
- Bastian, D. (2006). Modes of knowledge migration: Regional assimilation of knowledge and the politics of bringing knowledge into the region. *European Planning Studies*, 14(5), 601–619.
- Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): past, present, and future. *Scientometrics*, 52(3), 365–377.
- Beaverstock, J. V. (2004). Managing Across Borders: Knowledge Management and Expatriation in Professional Service Legal Firms. *Journal of Economic Geography*, 4, 157-179.
- Beine, M., Bertoli, S., & Moraga, J. F. (2014). A Practitioner's Guide to Gravity Models of International Migration.
- Beine, M., Docquier, F., & Rapoport, H. (2001). Brain drain and economic growth: theory and evidence. *Journal of Development Economics*, 64(1), 275–289.
- Beine, Michel, Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, 39(4), 496–512.
- Belderbos, R., Carree, M., & Lokshin, B. (2004). Cooperative R&D and firm performance. *Research Policy*, 33(10), 1477–1492.

- Bellini, E., Ottaviano, G. I. P., Pinelli, D., & Prarolo, G. (2013). Cultural diversity and economic performance: evidence from European regions. In *Geography, Institutions and Regional Economic Performance* (pp. 121–141). Springer.
- Belot, M. V., & Hatton, T. J. (2012). Immigrant Selection in the OECD. *The Scandinavian Journal of Economics*, 114(4), 1105–1128.
- Bergstrand, J. H. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, 474–481.
- Berry, A. R., & Soligo, R. (1969). Some Welfare Aspects of International Migration. *Journal of Political Economy*, 77(5), 778–794.
- Bertoli, S., Dequiedt, V., & Zenou, Y. (2016). Can selective immigration policies reduce migrants' quality? *Journal of Development Economics*, 119, 100–109.
- Bhagwati, J. N. (1975). The brain drain. *International Social Science Journal*, 28, 691–729.
- Bhagwati, J. N., & Hamada, K. (1974). The Brain-Drain, International Integration of Markets for Professionals and Unemployment. *Journal of Development Economics*, 1, 19–42.
- Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, 6(2), 173–189.
- Bjorkman, I., Barner-Rasmussen, W., & Li, L. (2004). Managing knowledge transfer in MNCs: The impact of headquarters control mechanisms. *Journal of International Business Studies*, 35(5), 443–455.
- Black, G. C., & Stephan, P. E. (2010). The economics of university science and the role of foreign graduate students and postdoctoral scholars. In *American universities in a global market* (pp. 129–161). University of Chicago Press.

- Bollard, A., McKenzie, D., Morten, M., & Rapoport, H. (2011, May). Remittances and the Brain Drain Revisited: The Microdata Show That More Educated Migrants Remit More. *The World Bank Economic Review*.
- Bonache, J., & Brewster, C. (2001). Knowledge transfer and the management of expatriation. *Thunderbird International Review*, 43(1), 145–168.
- Borjas, G. J. (2003). The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *Quarterly Journal of Economics*, 118(4), 1335–1374.
- Borjas, G. J. (2004). Do foreign students crowd out native students from graduate programs? National Bureau of Economic Research.
- Borjas, G. J. (2009). Immigration in High-Skill Labor Markets: The Impact of Foreign Students on the Earnings of Doctorates. In Freeman, R. B. & D. L. Goroff (eds.), *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*. University of Chicago Press.
- Borjas, G. J., & Doran, K. B. (2012). The Collapse of the Soviet Union and the Productivity of American Mathematicians*. *The Quarterly Journal of Economics*, 127(3), 1143–1203.
- Bozkurt, Ö. (2006). Wired for work: highly-skilled employment and global mobility in mobile telecommunications multinationals. In M. P. Smith and A. Favell (eds) *The human face of global mobility: international highly skilled migration in Europe, North America and the Asia-Pacific* (pp. 211–46). New Brunswick: Transaction Publishers.
- Breschi, S., & Lissoni, F. (2005). “Cross-firm” Inventors and Social Networks: Localized Knowledge Spillovers Revisited. *Annales d’Economie et de Statistique*, 189–209.

- Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localised knowledge flows. *Journal of Economic Geography*, 9(4), 439–468.
- Breschi, S., Lissoni, F., & Miguelez, E. (2015). Foreign inventors in the US: Testing for Diaspora and Brain Gain Effects. Paper Presented at the 8th International Conference on Migration and Development.
- Breschi, S., Lissoni, F., & Miguelez, E. (2017). Foreign-origin inventors in the USA: testing for diaspora and brain gain effects. *J Econ Geogr Lbw044*. Doi: 10.1093/Jeg/Lbw044.
- Breschi, S., Lissoni, F., & Noumedem, C. T. (2016). Migration and innovation: a survey of recent studies. *Handbook on the Geographies of Innovation*, 382.
- Breschi, S., Lissoni, F., & Tarasconi, G. (2014). Inventor Data for Research on Migration Innovation: A Survey and a Pilot. WIPO Economic Research Working Paper, 17, World Intellectual Property Organization –.
- Bresson, C. (1996). *C. Economic interdependence and innovative activity*. Cheltenham: Edward Elgar.
- Brinkerhoff, J. M. (2006). *Diaspora Mobilization Factors and Policy Options*, in “*Converting Migration Drains into Gains: Harnessing the Resources of Overseas Professionals*” (Wescott C. and J. Brinkerhoff). Asian Development Bank.
- Brown, M. (2002). Intellectual diaspora networks: their viability as a response to highly skilled migration. *Autrepart*, 22, 167–178.
- Burgers, J., & Touburg, J. (2013). International mobility of professional knowledge from the Global South: Indian IT workers in the Netherlands. *Global Networks*, 13(4), Article first published online: 12 FEB 2013.

- Burkert, C., Niebuhr, A., & Wapler, R. (2008). Regional Disparities in Employment of High-Skilled Foreigners – Determinants and Options for Migration Policy in Germany. *Journal of International Migration and Integration*, 9(4).
- Buttlar, L. (1999). Information sources in library and information science doctoral research. *Library & Information Science Research*, 21(2), 227–245.
- Chacko, E. (2007). From Brain Drain to Brain Gain: Reverse Migration to Bangalore and Hyderabad, India's Globalizing High Tech Cities. *GeoJournal*, 68(2–3), 131–140.
- Chaloff, J., & Lemaitre, G. (2009). Managing highly-skilled labour migration: a comparative analysis of migration policies and challenges in OECD countries (OCED Social, Employment and Migration WP). Paris: OECD Publishing.
- Champion, A. G. (1994). International Migration and Demographic Change in the World. *Urban Studies*, 13, 653–677.
- Chander, A. (2001). Diaspora Bonds. *New York University Law Review*, 76, 1005–1043.
- Chellaraj, G., Maskus, K. E., & Mattoo, A. (2008). The Contribution of International Graduate Students to US Innovation. *Review of International Economics*, 16(3), 444–462.
- Cincera, M., Kempen, L., Van Pottelsberghe, B., Veugelers, R., & Villegas, C. (2003). Productivity growth, R&D and the role of international collaborative agreements: Some evidence for Belgian manufacturing companies. *Brussels Economic Review*, 46(3), 107–140.
- Clarivate Analytics. (2017). Web of science. Trust the Difference. Web of Science Fact Book. Available Online: [Http://Images. Info. Science. Thomsonreuters. Biz/Web/ThomsonReutersScience/% 7Bd6b7faae-3cc2-4186-8985-A6ecc8cce1ee% 7D_Crv_WoS_Upsell_Factbook_A4_FA_LR_edits. Pdf](http://Images.Info.Science.Thomsonreuters.Biz/Web/ThomsonReutersScience/%7Bd6b7faae-3cc2-4186-8985-A6ecc8cce1ee%7D_Crv_WoS_Upsell_Factbook_A4_FA_LR_edits.Pdf) (Accessed on 10 October 2017).

- Clarke, B. L. (1967). Communication patterns of biomedical scientists. I. Multiple authorship and sponsorship of Federal Program volunteer papers. In *Federation Proceedings* (Vol. 26, pp. 1288–1292).
- Coe, D. T., & Helpman, E. (1995). International R&D spillovers. *European Economic Review*, 39(5), 859–887.
- Coe, D. T., & Moghadam, R. (1993). Capital and Trade as Engines of Growth in France: An Application of Johansen's Cointegration Methodology. IMF Working Paper 93/11, 40(3), 542–566.
- Coffano, M., & Tarasconi, G. (2014). CRIOS-Patstat database: sources, contents and access rules. Center for Research on Innovation, Organization and Strategy, CRIOS Working Paper, 1(1).
- Collings, D. G., & Scullion, H. (2009). Global staffing. *The International Journal of Human Resource Management*, 20(6), 1249–52.
- Conchi, S., & Michels, C. (2014). Scientific mobility – An analysis of Germany, Austria, France and Great Britain. Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis No. 41, Karlsruhe, March 2014.
- David, P. A. (1993). Intellectual property institutions and the panda's thumb: Patents, copyrights, and trade secrets in economic theory and history. National Academy.
- De Backer, K., & Basri, E. (2008). The internationalisation of R&D. *Staying Competitive in the Global Economy: Compendium of Studies on Global Value Chains*, 219–248.
- De Filippo, D., Casado, E. S., & Gómez, I. (2009). Quantitative and qualitative approaches to the study of mobility and scientific performance: a case study of a Spanish university. *Research Evaluation*, 18(3), 191–200.
- De Haas, H. (2010). Migration and development: A theoretical perspective 1. *International Migration Review*, 44(1), 227–264.

- Decramer, A., Goeminne, S., & Smolders, C. (2013). The impact of internationalization on volume and quality of scholarly publication performance. *Public Money & Management*, 33(2), 111–117.
- Docquier, F., Lohest, O., & Marfouk, A. (2007). Brain drain in developing countries. *World Bank Economic Review*, 21(2), 193–218.
- Docquier, F., Lowell, B. L., & Marfouk, A. (2009). A gendered assessment of highly skilled emigration. *Population and Development Review*, 35(2), 297–321.
- Docquier, F., & Rapoport, H. (2012). Globalization, Brain Drain, and Development. *Journal of Economic Literature*, 50(3), 681–730.
- Dosi, G. (1988). Sources, procedures and microeconomic effects of innovation. *Journal of Economic Literature*, 36, 1126–1171.
- Dunlevy, J. A. (2006). The influence of corruption and language on the protrade effect of immigrants: Evidence from the American States. *Review of Economics and Statistics*, 88(1), 182–186.
- Dustmann, C., & Görlach, J.-S. (2016). The economics of temporary migrations. *Journal of Economic Literature*, 54(1), 98–136.
- Dustmann, C., & Weiss, Y. (2007). Return migration: theory and empirical evidence from the UK. *British Journal of Industrial Relations*, 45(2), 236–256.
- Eaton, J., & Kortum, S. (1996). Trade in Ideas: Patenting and Productivity in the OECD. *Journal of International Economics*, 40(3–4), 251–278.
- Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415–430.
- Faems, D., Van Looy, B., & Debackere, K. (2005). Interorganizational collaboration and innovation: Toward a portfolio approach. *Journal of Product Innovation Management*, 22(3), 238–250.

- Felbermayr, G. J., Jung, B., & Toubal, F. (2010). Ethnic networks, information, and international trade: Revisiting the evidence. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 41–70.
- Felbermayr, G. J., & Toubal, F. (2012). Revisiting the trade-migration nexus: Evidence from new OECD data. *World Development*, 40(5), 928–937.
- Finn, M. (2010). Stay rates of foreign doctorate recipients from US universities, 2005 (No. 2007 (No. 10-SEP-0168)). Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN.
- Foley, C. F., & Kerr, W. R. (2013). Ethnic innovation and US multinational firm activity. *Management Science*, 59(7).
- Franzoni, C., Scellato, G., & Stephan, P. (2012). Foreign Born Scientists: Mobility Patterns for Sixteen Countries. National Bureau of Economic Research Working Paper Series, No. 18067. Retrieved from <http://www.nber.org/papers/w18067>
- Freeman, R. B. (2013). Globalization of scientific and engineering talent: international mobility of students, workers, and ideas and the world economy. *Economics of Innovation and New Technology*, 19(5), 393–406.
<https://doi.org/10.1080/10438590903432871>
- Gaillard, A. M., & Gaillard, J. (1997). The international mobility of brains: Exodus or circulation? *Science, Technology & Society*, 2(2), 195–228.
- Galbraith, J. R., & Edstrom, A. (1976). International Transfer of Managers: Some Important Policy Considerations. *Columbia Journal of World Business*, 11(2), 100–112.
- Gibson, J., & McKenzie, D. (2012). The Economic Consequences of ‘Brain Drain’ of the Best and Brightest: Microeconomic Evidence from Five Countries*. *The Economic Journal*, 122(560), 339–375.

- Gibson, J., & McKenzie, D. (2014). Scientific mobility and knowledge networks in high emigration countries: Evidence from the Pacific. *Research Policy*, 43(9), 1486–1495.
- Gill, B. (2005). Homeward bound? The experience of return mobility for Italian scientists. *Innovation*, 18, 319–341.
- Giovannetti, G., & Lanati, M. (2015). The indirect pro-trade effects of Indian ethnic networks. DEMO-India Research Report 2015/14.
- Girma, S., & Yu, Z. (2000). The link between immigration and trade: evidence from the UK. Centre for the Research on Globalisation and Labour Markets (GLM). GEP Research Papers, (23).
- Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research* (pp. 257–276). Springer.
- Glänzel, W., Schubert, A., & Czerwon, H.-J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, 45(2), 185–202.
- Gould, D. M. (1994). Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows. *The Review of Economics and Statistics*, 76(2), 302–316.
- Granstrand, O., Hakanson, L., & Sjölander, S. (1993). Internationalization of R&D—a survey of some recent research. *Research Policy*, 22(5–6), 413–430.
- Greve, P., Nielsen, S., & Ruigrok, W. (2009). Transcending borders with international top management teams: A study of European financial multinational corporations. European School of Management, University of Glasgow, 27, 214–215, (accessed February 27, 2010).
- Griliches, Z. (1988). Productivity Puzzles and R&D: Another Nonexplanation. 2, 4, 9–21.
- Grubel, H. G., & Scott, A. D. (1966). The Immigration of Scientists and Engineers to the United States, 1949-61. *Journal of Political Economy*, 74(4), 368–378.

- Guellec, D., & de la Potterie, B. van P. (2001). The internationalisation of technology analysed with patent data. *Research Policy*, 30(8), 1253–1266.
- Hall, B., Jaffe, A., & Trajtenberg, M. (2001). The NBER patent citations data file: Lessons, insights and methodological tools (NBER working paper no. 8498).
- Harzing, A. W. (2001). An Analysis of the Functions of International Transfer of Managers in MNCs. *Employee Relations*, 23, 581–598.
- Head, K., & Ries, J. (1998). Immigration and trade creation: econometric evidence from Canada. *Canadian Journal of Economics*, 47–62.
- Heijltjes, M. G., Olie, R., & Glunk, U. (2003). Internationalization of Top Management Teams in Europe. *European Management Journal*, 21, 89–97.
- Herander, M. G., & Saavedra, L. A. (2005). Exports and the structure of immigrant-based networks: the role of geographic proximity. *Review of Economics and Statistics*, 87(2), 323–335.
- Hocking, J. B., Brown, M., & Harzing, A. (2004). A knowledge transfer perspective of strategic assignment purposes and their path-dependent outcomes. *Journal of Human Resource Management*, 15, 565–586.
- Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5), 662–673.
- Hornung, E. (2014). Immigration and the diffusion of technology: The Huguenot diaspora in Prussia. *American Economic Review*, 104(1), 84–122.
- Hugo, G. (2013). What we know about circular migration and enhanced mobility. *Migration Policy Institute*, 7, 1–10.
- Hunt, J. (2009). Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa (National Bureau of Economic Research, Inc, NBER Working Papers:

14920. 2009.). Retrieved from
<http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=econ&AN=1033983>
- Hunt, J. (2013). Are Immigrants the Best and Brightest U.S. Engineers? National Bureau of Economic Research Working Paper Series, No. 18696. Retrieved from
<http://www.nber.org/papers/w18696>
- Hunt, J., & Gauthier-Loiselle, M. (2010). How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics*, 2(2), 31–56.
- Ikiara, M. M. (2003). Foreign Direct Investment (FDI), Technology Transfer and Poverty Alleviation: Africa's Hopes and Dilemma. African Technology Policy Studies Network (ATPS) Special Paper Series No. 16.
- Imbens, G., & Wooldridge, J. (2007). Difference-in-differences estimation. National Bureau of Economics Research Working Paper.
- Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *The American Economic Review*, 76, 984–1001.
- Jaffe, A. B., & Caballero, R. J. (1993). How high are the giants' shoulders: an empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. In Blanchard, Olivier, Fischer, Stanley (Eds.) (NBER Macroeconomics Annual, Vol. 8, pp. 15–86). MIT Press.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3), 577–598.
- Javorcik, B. S., Ozden, C., Spatareanu, M., & Neagu, C. (2011). Migrant networks and foreign direct investment. *Journal of Development Economics*, 94(2), 231–241.

- Jin, B., Rousseau, R., Suttmeier, R. P., & Cao, C. (2007). The role of ethnic ties in international collaboration: the overseas Chinese phenomenon.
- Johnson, H. G. (1967). Some Economic Aspects of Brain Drain. *The Pakistan Development Review*, 7(3), 379–411.
- Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, 322(5905), 1259–1262.
- Jonkers, K., & Cruz-Castro, L. (2013). Research upon return: The Effect of International Mobility on Scientific Ties, Production and Impact. *Research Policy*, 42, 1366– 1377.
- Jonkers, K., & Tijssen, R. (2008). Chinese Researchers Returning Home: Impacts of International Mobility on Research Collaboration and Scientific Productivity. *Scientometrics*, 77(2), 309–333.
- Jonkers, Koen. (2008). A comparative study of return migration policies targeting the highly skilled in four major sending countries.
- Jonkers, Koen. (2009). Scientific mobility and international research collaboration: Case of China. *On Knowledge Systems*, 78.
- Jöns, H. (2009). ‘Brain circulation’ and transnational knowledge networks: studying long-term effects of academic mobility to Germany, 1954–2000. *Global Networks*, 9(3), 315–338.
- Kahn, S., & MacGarvie, M. (2011). The effects of the foreign Fulbright program on knowledge creation in science and engineering. In *The rate and direction of inventive activity revisited* (pp. 161–197). University of Chicago Press.
- Kamalski, J., & Plume, A. (2013). Comparative Benchmarking of European and US Research Collaboration and Researcher Mobility. A Report Prepared in Collaboration between *Science Europe and Elsevier’s SciVal Analytics, September 2013*.

- Kangasniemi, M., Winters, L. A., & Commander, S. (2007). Is the Medical Brain Drain Beneficial? Evidence from Overseas Doctors in the U.K. *Social Science & Medicine*, 65(5), 915–923.
- Kapur, D. (2001). Diasporas and Technology Transfer. *Journal of Human Development*, 2(2).
- Kapur, D., & McHale, J. (2005). Give us your best and brightest: The global hunt for talent and its impact on the developing world. Brookings Inst Press.
- Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31–43.
- Keller, W. (1997). Are international R&D spillovers trade-related?: Analyzing spillovers among randomly matched trade partners. *European Economic Review*, 42(8), 1469–1481.
- Keller, W. (2001, March). The Geography and Channels of Diffusion at the World's Technology Frontier. NBER Working Paper No. 8150.
- Kenney, M., Breznitz, D., & Murphree, M. (2013). Coming back home after the sun rises: Returnee entrepreneurs and growth of high tech industries. *Research Policy*, 42(2), 391–407. <https://doi.org/10.1016/j.respol.2012.08.001>
- Kerr, S. P., Kerr, W., Ozden, C., & Parsons, C. R. (2016). Global talent flows. *The Journal of Economic Perspectives*, 30(4), 83–106.
- Kerr, W. R. (2007). The Ethnic Composition of US Inventors (Harvard Business School, Working Papers No. 08–006) (p. 27 pages). Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=econ&AN=0971283>
- Kerr, W. R. (2008). Ethnic Scientific Communities and International Technology Diffusion. *Review of Economics and Statistics*, 90(3), 518–537. <https://doi.org/10.1162/rest.90.3.518>

- Kerr, W. R. (2009). The Agglomeration of US Ethnic Inventors (NBER Working Papers No. 09-003). NBER Working Papers. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=econ&AN=0989119>
- Klein, P., & Ventura, G. J. (2007). TFP differences and the aggregate effects of labor mobility in the long run. *The BE Journal of Macroeconomics*, 7(1).
- Kopp, R. (1994). International Human Resource Policies and Practices in Japanese European, and United States Multinationals. *Human Resource Management Journal*, 33, 581–599.
- Koser, K., & Salt, J. (1997). The Geography of Highly Skilled International Migration. *International Journal of Population Geography*, 3(4), 285–303.
- Kugler, M., & Rapoport, H. (2007). International labor and capital flows: Complements or substitutes? *Economics Letters*, 94(2), 155–162.
- Kuznetsov, Y. (Ed.). (2006). *Diaspora networks and the international migration of skills: how countries can draw on their talent abroad*. World Bank Publications, Washington, DC.
- Lata, R., Scherngell, T., & Brenner, T. (2012). Integration Processes in European R&D: A comparative spatial interaction approach using project based R&D networks, co-publication networks and co-patent networks.
- Laudel, G. (2003). Studying the brain drain: Can bibliometric methods help? *Scientometrics*, 57(2), 215–237.
- Leonard, D., & Sensiper, S. (1998). The role of tacit knowledge in group innovation. *California Management Review*, 40, 112–132.
- Lodigiani, E. (2009). Diaspora externalities and technology diffusion. *Economie Internationale*, (3), 43–64.

- Lowell, B. L., & Findlay, A. (2001). Migration of highly skilled persons from developing countries: impact and policy responses. *International Migration Papers*, 44(25).
- Lowendahl, B., Revang, O., & Fosstenlokken, S. N. (2001). Knowledge and Value Creation in Professional Service Firms. *Human Relations*, 54, 911–931.
- Lucas, R. E. (1990). Why Doesn't Capital Flow from Rich to Poor Countries? *The American Economic Review*, 80(2 (P&P)), 92–96.
- Luukkonen, T., Tijssen, R. J. W., Persson, O., & Siversten, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28, 15–36.
- MacGarvie, M. (2005). The determinants of international knowledge diffusion as measured by patent citations. *Economics Letters*, 87(1), 121–126.
- MacGarvie, M. (2006). Do Firms Learn from International Trade? *The Review of Economics and Statistics*, 88(1), 46–60.
- Maggioni, M. A., Nosvelli, M., & Uberti, T. E. (2007). Space versus networks in the geography of innovation: A European analysis. *Papers in Regional Science*, 86(3), 471–493.
- Mahroum, S. (2005). The international policies of brain gain: A review. *Technology Analysis & Strategic Management*, 17(2), 219–230.
- Mansfield, E., & Romeo, A. (1980). Technology transfer to overseas subsidiaries by U.S.-based firms. *The Quarterly Journal of Economics*, 95, 737–750.
- Martins, J. M. (2011). Knowledge Transfer to the Subsidiaries: A Case Study of Multinational Enterprise Operating in Overseas. *Review of Economics & Finance*.
- Maskus, K. E., & Fink, C. (2005). Intellectual property and development: lessons from recent economic research. *The World Bank*.

- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, E. (1993). Theories of international migration: A review and appraisal. *Population and Development Review*, 431–466.
- Massey, D. S., & Parrado, E. (1994). Migradollars: The remittances and savings of Mexican migrants to the USA. *Population Research and Policy Review*, 13(1), 3–30.
- Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, 23(4), 1249–1274.
- Mayer, T., & Zignago, S. (2011). Notes on CEPII's Distances Measures: the GeoDist Database. CEPII Working Paper 2011-25.
- McAusland, C., & Kuhn, P. (2011). Bidding for brains: Intellectual property rights and the international migration of knowledge workers. *Journal of Development Economics*, 95(1), 77–87. <https://doi.org/10.1016/j.jdeveco.2010.02.002>
- Meijering, L., & van Hoven, B. (2003). Imagining difference: the experiences of 'transnational' Indian IT professionals in Germany. *Area*, 35(2), Article first published online: 4 JUN 2003.
- Merton, R. K. (1968). *Social theory and social structure*. Simon and Schuster.
- Meyer, J. P., & Wattiaux, J. P. (2006). Diaspora Knowledge Networks: Vanishing Doubts, Increasing Evidence, in: *International Journal on Multicultural Societies*. *IJMS*, 8(1), 4–24.
- Meyer, J.-B. (2001). Network Approach versus Brain Drain: Lessons from the Diaspora. *International Migration*, 39(5), 91–110. <https://doi.org/10.1111/1468-2435.00173>
- Migueluez, E. (2016). Inventor Diasporas and the Internationalization of Technology. *The World Bank Economic Review*, Lhw013.

- Migueluez, E., & Fink, C. (2013). Measuring the International Mobility of Inventors: A New Database. World Intellectual Property Organization-Economics and Statistics Division.
- Migueluez, E., & Noumedem, C. T. (2017). Immigration externalities, knowledge flows and brain gain. Groupe de Recherche en Economie Théorique et Appliquée.
- Moed, H. F., & Halevi, G. (2014). A bibliometric approach to tracking international scientific migration. *Scientometrics*, 101(3), 1987–2001.
- Moed, H. F., & Plume, A. (2013). Studying scientific migration in Scopus. *Scientometrics*, 94(3), 929–942.
- Morgan, G. (2001). Transnational Communities and Business Systems. *Global Networks*, 1, 113–130.
- Moser, P., Voena, A., & Waldinger, F. (2014). German Jewish émigrés and US invention. *American Economic Review*, 104(10), 3222–55.
- Moses, J. W., & Letnes, B. (2004). The economic costs to international labor restrictions: Revisiting the empirical discussion. *World Development*, 32(10), 1609–1626.
- Mountford, A. (1997). Can a brain drain be good for growth in the source economy? *Journal of Development Economics*, 53(2), 287–303.
- Mukherjee, S., Romero, D. M., Jones, B., & Uzzi, B. (2017). The nearly universal link between the age of past knowledge and tomorrow’s breakthroughs in science and technology: The hotspot. *Science Advances*, 3(4), e1601315.
- Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics* 21, 3, 313–323.
- Nathan, M. (2015). Same difference? Minority ethnic inventors, diversity and innovation in the UK. *Journal of Economic Geography*, 15(1), 129–168.

- Nelson, R. R., & Phelps, E. S. (1966). Investment in Humans, Technological Diffusion, and Economic Growth. *The American Economic Review*, 56(1/2 (March 1966)), 66–75.
- Niebuhr, A. (2010a). Migration and Innovation: Does Cultural Diversity Matter for Regional R&D Activity? *Papers in Regional Science*, 89(3), 563–585.
- Nkiko, C., & Adetoro, N. (2007). Pioneer bachelor degree: Citation analysis of Covenant University students' research projects. *Library Philosophy and Practice*.
- No, Y., & Walsh, J. P. (2010). The importance of foreign-born talent for US innovation. *Nature Biotechnology*, 28(3), 289–291.
- Nooteboom, B. (1999). Innovation and inter-firm linkages: new implications for policy. *Research Policy*, 28(8), 793–805.
- OECD. (2008). *The global competition for talent: Mobility of the highly skilled*. OECD Paris.
- Okubo, Y., & Zitt, M. (2004). Searching for research integration across Europe: a closer look at international and inter-regional collaboration in France. *Science and Public Policy*, 31(3), 213–226.
- Ortega, F., & Peri, G. (2009). The causes and effects of international mobility: Evidence from OECD countries 1980- 2005. NBER Working Paper, 14833.
- Ortega, F., & Peri, G. (2013). The Role of Income and Immigration Policies in Attracting International Migrants. *Migration Studies*, 1(1), 47–74.
- Ottaviano, G. I. P., & Peri, G. (2006). The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography*, 6(1), 9–44.
- Özden, Ç., Parsons, C. R., Schiff, M., & Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, 25(1), 12–56.
- Ozgen, C., Nijkamp, P., & Poot, J. (2011). Immigration and innovation in European regions. Discussion paper series//Forschungsinstitut zur Zukunft der Arbeit.

- Ozgen, C., Nijkamp, P., & Poot, J. (2013). The impact of cultural diversity on firm innovation: evidence from dutch micro-data. *IZA Journal of Migration*, 2(18).
- Peixoto, J. (1999). *A Mobilidade Internacional dos Quadros - Migrações Internacionais, Quadros e Empresas Transnacionais em Portugal*. Oeiras: Celta Editora.
- Peixoto, J. (2006). The International Mobility of Highly Skilled Workers in Transnational Corporations: The Macro and Micro Factors of the Organizational Migration of Cadres. *International Migration Review*, 35(4), Article first published online: 23 FEB 2006.
- Perkins, S. J. (1997). *Internationalization: The People Business*. London: Kogan Page.
- Pierson, A. S., & Cotgreave, P. (2000). Citation figures suggest that the UK brain drain is a genuine problem. *Nature*, 407(6800).
- Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86(3), 423–443.
- Pyka, A. (1997). Informal Networking. *Telenovation*, 17(4), 207–220.
- Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics*, 48(1), 7–35.
- Rauch, J. E., & Trindade, V. (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1), 116–130.
- Ravenstein, E. G. (1889). The Laws of Migration. *Journal of the Royal Statistical Society*, 52(2), 241–305.
- Removille, J., & Clarysse, B. (1999). Intra-European scientific co-operation: measuring policy impact. *Research Evaluation*, 8(2), 99–109.
- Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science*, 59(2), 504–525.

- Rodrigue, J. P., Comtois, C., & Slack, B. (2009). *The Geography of Transport Systems*. London, New York: Routledge. ISBN 0-415-88324-7.
- Rodriguez, A. R. (2002). Determining factors in entry choice for international expansion. The case of the Spanish hotel industry. *Tourism Management*, 23(6), 598–607.
- Roessner, J. D. (2002). Outcome measurement in the USA: state of the art. *Research Evaluation*, 11(2), 85–93.
- Romer, P. M. (1996). Increasing Returns and Long run Growth. *Journal of Political Economy*, 94, 1002–1037.
- Roos, D. H., Calabró, L., de Jesus, S. L., Souza, D. O., Barbosa, N. V., & da Rocha, J. B. T. (2014). Brazilian scientific production in areas of biological sciences: a comparative study on the modalities of full doctorate in Brazil or abroad. *Scientometrics*, 98(1), 415–427.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38.
- Salt, J. (2008). Managing new migrations in Europe: Concept and reality in the ICT sector. In Bonifazi, C. et al. (eds) (2008) *International Migration in Europe. New Trends and New Methods of Analysis* (pp. 19–36). IMSCOE Research, Amsterdam University Press.
- Salt, J., & Wood, P. (2012). Recession and International Corporation Mobility. *Global Networks*, 12(4), 425–445.
- Sanna-Randaccio, F., & Veugelers, R. (2007). Multinational Knowledge Spillovers with Decentralised R&D: A Game-Theoretic Approach. *Journal of International Business Studies*, 38, 47–63.

- Santos-Silva, J. M. C., & Tenreyro, S. (2006). The Log of Gravity. *Review of Economics and Statistics*, 88(4), 641–658.
- Saxenian, A. (1999). *Silicon Valley's New Immigrant Entrepreneurs*. San Francisco, CA: Public Policy Institute of California.
- Saxenian, A., Motoyama, Y., & Quan, X. (2002). Local and global networks of immigrant professionals in Silicon Valley. *Public Policy Instit. of CA*.
- Scellato, G., Franzoni, C., & Stephan, P. (2015). Migrant scientists and international networks. *Research Policy*, 44(1), 108–120.
- Shin, J. C., Jung, J., Postiglione, G. A., & Azman, N. (2014). Research productivity of returnees from study abroad in Korea, Hong Kong, and Malaysia. *Minerva*, 52(4), 467–487.
- Song, H. (1997). From brain drain to reverse brain drain: Three decades of Korean experience. *Science, Technology and Society*, 2(2), 317–345.
- Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). Measuring patent quality. *OECD Science, Technology and Industry Working Papers*, No. 2013/03.
- Stark, O., Helmenstein, C., & Prskawetz, A. (1997). A brain gain with a brain drain. *Economics Letters*, 55(2), 227–234.
- Stark, O., Helmenstein, C., & Prskawetz, A. (1998). Human capital depletion, human capital formation, and migration: a blessing or a “curse”? *Economics Letters*, 60(3), 363–367.
- Stark, O., & Wang, Y. (2002). Inducing human capital formation: migration as a substitute for subsidies. *Journal of Public Economics*, 86(1), 29–46.
- Stephan, P. E. (2010). The economics of science. In *Handbook of the Economics of Innovation* (Vol. 1, pp. 217–273). Elsevier.
- Stephan, P. E. (2012). *How economics shapes science* (Vol. 1). Harvard University Press
Cambridge, MA.

- Stephan, P. E., & Levin, S. G. (2001). Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review*, 20(1), 59–79.
- Straubhaar, T. (2000). International mobility of the highly skilled: Brain gain, brain drain or brain exchange. HWWA Discussion Paper, No. 88.
- Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs. *Best Practices in Quantitative Methods*, 155–176.
- Stuen, E. T., Mobarak, A. M., & Maskus, K. E. (2012). Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in US Doctoral Programmes*. *The Economic Journal*, 122(565), 1143–1176. <https://doi.org/10.1111/j.1468-0297.2012.02543.x>
- Su, X. (2012). International Doctoral Science and Engineering Students: Impact on Cohorts' Career Prospects. *Journal of Studies in International Education*. Retrieved from <http://jsi.sagepub.com/content/early/2012/12/05/1028315312467354.abstract>
- Tagliacozzo, R. (1977). Self-citations in scientific literature. *Journal of Documentation*, 33(4), 251–265.
- Tan, D., & Mahoney, J. T. (2006). Why a Multinational Firm Chooses Expatriates Integrating Resource-Based, Agency and Transaction Costs Perspectives. *Journal of Management Studies*, 43(3), 458–484.
- Taylor, E. (1999). The new economics of labour migration and the role of remittances in the migration process. *International Migration*, 37(1), 63–88.
- Thomas, D. C. (2008). *Cross-cultural management: Essential concepts* (2nd ed.). CA: Sage: Thousand Oaks.
- Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, 95(1), 450–460.
- Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund.

- Trippel, M. (2013). Scientific mobility and knowledge transfer at the interregional and intraregional level. *Regional Studies*, 47(10), 1653–1667.
- Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth*, 11(2), 97–127.
- Velema, T. A. (2012). The Contingent Nature of Brain Gain and Brain Circulation: their Foreign Context and the Impact of Return Scientists on the Scientific Community in their Country of Origin. *Scientometrics*, 93(3).
- Von Hippel, E. (1988). *The Sources of Innovation*. Nueva York: Oxford University Press.
- Wadhwa, V., Jasso, G., Rissing, B., Gereffi, G., & Freeman, R. (2007). Intellectual Property, the Immigration Backlog, and a Reverse Brain-Drain: America's New Immigrant Entrepreneurs, Part III. Part III (August 22, 2007).
- Wadhwa, V., Rissing, B., Saxenian, A., & Gereffi, G. (2007). Education, Entrepreneurship and Immigration: America's New Immigrant Entrepreneurs, Part II. Part II (June 11, 2007).
- Wadhwa, V., Saxenian, A., Freeman, R. B., & Gereffi, G. (2009). America's loss is the world's gain: America's new immigrant entrepreneurs, Part 4.
- Wadhwa, V., Saxenian, A., Freeman, R. B., & Salkever, A. (2009). Losing the world's best and brightest: America's new immigrant entrepreneurs, part V.
- Wadhwa, V., Saxenian, A., Rissing, B., & Gereffi, G. (2007). America's new Immigrant entrepreneurs: Part I. *Duke Science, Technology & Innovation Paper*, (23).
- Wagner, C. S. (2005). Six case studies of international collaboration in science. *Scientometrics*, 62(1), 3–26.
- Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10), 1608–1618.

- Webb, C., Denis, H., Harhoff, D., & Hoisl, K. (2005). Analysing European And International Patent Citations: A Set of EPO Patent Database Building Blocks. STI Working Paper 2005/9.
- Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1), 117–131.
- Widmaier, S., & Dumont, J.-C. (2011). Are recent immigrants different? A new profile of immigrants in the OECD based on DIOC 2005/06. Paris: OECD Publishing.
- Windmeijer, F. A., & Santos Silva, J. M. (1997). Endogeneity in count data models: an application to demand for health care. *Journal of Applied Econometrics*, 12(3), 281–294.
- World Bank. (2018). Moving for Prosperity: Global Migration and Labor Markets. Policy Research Report. Washington (Policy Research Report). Washington, DC: World Bank. Retrieved from doi:10.1596/978-1-4648-1281-1.
- Yoshikane, F., & Kageura, K. (2004). Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics* 60, 3, 435–446.
- Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, 35(1), 83–104.