Université Clermont Auvergne
École Doctorale des Sciences Économiques, Juridiques, et de Gestion
Centre d'Études et de Recherches sur le Développement International
(CERDI)

THREE ESSAYS ON INTERNATIONAL MIGRATION

Thèse nouveau Régime
Présentée et soutenue publiquement le 14 Septembre 2017
Pour l'obtention du titre de Docteur *ès* Sciences Économiques

Par
Jérôme VALETTE

Sous la direction de :
M. Simone BERTOLI et M. Jean-Louis COMBES

Membres du jury

Olivier BARGAIN	Professeur, Université de Bordeaux (Rapporteur)
Simone BERTOLI	Professeur, Université Clermont Auvergne (Directeur)
Isabelle CHORT	Professeur, Université de Pau et des Pays de l'Adour (Rapporteur)
Jean-Louis COMBES	Professeur, Université Clermont Auvergne (Directeur)
Vianney DEQUIEDT	Professeur, Université Clermont Auvergne (Suffragant)
Hillel RAPOPORT	Professeur, Université Paris 1 Panthéon-Sorbonne (Rapporteur)

À Anna, qui, sans jamais que nous le sachions tous deux, fut le but de tout cela.

Remerciements

Il est des phrases qui sont plus faciles à écrire que d'autres, des mots qui tombent sous l'évidence, et certains, que notre timidité naturelle nous ferait garder pour soi. L'exercice des remerciements qui précède chaque thèse vient naturellement balayer tout cela. Ainsi, à l'heure où je rédige ces quelques lignes, sachez que tout ce que je voudrais dire, à l'ensemble des personnes qui m'ont accompagné dans cette aventure, ne saurait être justement retranscrit ici, tant il faudrait de mots pour vous exprimer toute ma reconnaissance.

Mes premiers mots vont évidemment à mes deux directeurs, M. Bertoli Simone et M. Combes Jean-Louis. Sans votre aide, il va sans dire que la qualité de mes travaux n'aurait pas été la même. Au delà de vos précieux conseils, je voudrais avant tout vous remercier pour vos encouragements, votre dévouement et votre bienveillance, tout au long de ces trois années. Je crois avoir énormément appris à vos cotés, bien sûr, concernant la théorie économique et ses applications, mais plus encore, sur la rigueur que nécessite tout travail scientifique. Alors que cette thèse prend fin je mesure l'immense chance qui a été la mienne d'avoir pu bénéficier d'un encadrement de cette qualité et vous en suis profondément reconnaissant.

Je voudrais aussi adresser mes sincères remerciements à Olivier Bargain, Isabelle Chort, Vianney Dequiedt et Hillel Rapoport pour avoir non seulement accepté de relire les travaux présentés dans cette thèse mais pour y avoir apporté de précieux commentaires.

Un remerciement tout particulier va à mes co-auteurs, Frédéric, Sekou, Riccardo et Valantis. La co-écriture aura, pour sûr, été une des expériences les plus enrichissantes de cette thèse. Au delà de toutes les discussions que nous avons pu partager et de tout ce que vous m'avez appris, je voudrais souligner ici l'amitié dont vous avez toujours fait preuve à mon égard et qui m'est si chère aujourd'hui. J'espère que nos routes continueront de se croiser à l'avenir, le temps d'une bière, d'un tennis, d'une ballade à vélo ou d'un nouveau papier évidemment !

Aussi, je voudrais remercier mes compagnons de route Hippolyte, Ababacar et Victor qui m'ont entouré durant ces trois années et avec lesquels j'ai tant partagé.

Mention spéciale à Victor qui a du supporter une année de plus que les autres mes questionnements et mes digressions trop nombreuses ! Ton soutien quotidien et ton amitié durant cette dernière année de thèse sont d'une valeur inestimable.

J'ai bien sûr aussi une pensée pour l'ensemble des professeurs et du personnel du CERDI, et tout particulièrement pour Agnès, Chantal, Johan et Marie. Votre bonne humeur indéfectible et votre gentillesse sont des contributions immenses aux conditions de travail épanouissantes qu'offre le CERDI à ses doctorants.

Si cette thèse vient clôturer 3 années de doctorat au CERDI, l'aventure Clermontoise a démarré depuis 8 ans déjà et je voudrais ici remercier toute les personnes qui ont compté durant ces bien trop longues années d'études. À, Antoine et Victor, qui, plus que des amis ont été de vrais pères spirituels. Merci (Sidi) de nous avoir pris sous votre aile avec P.O. depuis notre premier jour en MAG1. Je n'oublierai évidemment pas de mentionner les nombreuses défaites (toujours avec dignité), les victoires (si si il y en a eu ...) et tout ces moments BDS et autres que nous avons pu partager. La création de la Moundialette restera à jamais la plus belle réalisation de ces années. Ne jamais l'avoir gagné, mon plus grand échec ! À Émilie, pour m'avoir enfin donné une bonne raison de descendre au 1$^{\text{er}}$ étage et à Laurène dont je n'oublierai assurément pas les paroles. À mes amis de toujours Hugo, Gaillou, J.B., P.O., Poussin, Sam et Vivi. A tous nos souvenirs déjà et ceux qu'il nous reste à construire ensembles. Je ne vous échangerais pour rien au monde.

Enfin, je voudrais terminer ici par remercier ma famille, ainsi que ma belle-famille, qui m'ont toujours soutenu et sans qui tout cela n'aurait que peu de sens. À mes parents et ma sœur je voudrais aussi dire qu'à l'heure de fonder ma propre famille je mesure combien grandir entouré d'autant d'amour est une chance. Je ne saurais convenablement vous remercier et vous exprimer ici toute mon affection. Et pour finir, à vous, Elodie et Anna. Vous qui êtes tout. Vous voir tous les jours sourire suffit à mon bonheur.

Résumé

Le sujet des migrations internationales a récemment fait l'objet d'une attention sans précédent dans l'opinion publique comme dans les médias. Or, si le débat sur les effets des migrations internationales semble plus que jamais d'actualité, celui-ci attire l'attention des économistes depuis plusieurs décennies déjà. La présente thèse s'insère ainsi dans la littérature économique sur les effets des migrations internationales en proposant trois essais empiriques sur les implications de la mobilité humaine, à la fois pour les migrants, les natifs dans leur pays d'accueil et leurs proches restés dans leur pays d'origine.

Le Chapitre 2 revisite de manière empirique l'impact du multiculturalisme (mesuré par un indice de diversité à l'intérieur du groupe des migrants et par des effets de contamination) sur les performances macro-économiques des États Américains sur la période 1960-2010. Nous distinguons les effets du multiculturalisme par niveaux d'éducation, en contrôlant pour les variables standards de la littérature sur la croissance ainsi que pour l'hétérogénéité inobservée et en prenant en compte le statut légal des migrants ainsi que leur âge d'entrée aux États-Unis. Dans le but d'identifier un effet causal, nous comparons plusieurs stratégies d'identification différentes de la littérature existante. Nos résultats convergent vers un effet robuste positif et significatif de la diversité des diplômés du tertiaire sur le PIB par tête. Aucun effet de la diversité pour les niveaux d'éducation inférieurs, ou d'effets de contamination ne sont mis en évidence.

Le Chapitre 3 s'insère dans la littérature sur les déterminants de la performance des migrants sur le marché du travail dans leur pays d'accueil. Nous regardons si l'attitude des natifs affecte ou non les durées de chômage des migrants en Allemagne. En utilisant des données de panel (GSOEP) au niveau individuel sur la période 1984-2012 et un modèle de durée, nous trouvons que des niveaux de confiance plus faibles des natifs envers les résidents d'un pays donné (mesurés à l'aide des enquêtes Eurobarometers) sont associés à des durées de chômage plus longues pour les immigrés originaires de ce pays. Nos résultats soulignent le fait que, différents groupes d'immigrés font face à des obstacles différents en fonction de leur origine, pour s'insérer sur le marché du

travail.

Le Chapitre 4 cherche quant à lui à comprendre si les migrants au niveau international contribuent ou non au progrès technologique dans les pays en développement en induisant un transfert de connaissances productives de leur pays d'accueil vers leur pays d'origine. En utilisant un indicateur pour le niveau de connaissances productives de chaque pays (ECI) et les stocks bilatéraux de migrants vers 20 pays de l'OCDE, nous montrons que la migration internationale est un canal de transmission important de la technologie.

Mots clés : Économie du Développement, Migrations Internationales, Diversité, Croissance, Durées de Chômage, Discriminations, Changements Technologiques.
Codes JEL : F63, F22, J61, J64, J71, 033.

Summary

International migration recently attracted unprecedented public attention and media coverage. However, while the debate on the effects on international migration on the economy seems now more relevant than ever, it already attracts the attention of economic researchers for decades. The present thesis provides three empirical studies that investigate the implications of international migration both for migrants themselves, natives in their host countries and those left behind.

Chapter 2 empirically revisits the impact of multiculturalism on the macroeconomic performance of US states over the 1960-2010 period. We test for skill-specific effects of multiculturalism, controlling for standard growth regressors and a variety of fixed effects, and accounting for the age of entry and legal status of immigrants. To identify causation, we compare various instrumentation strategies used in the existing literature. We provide converging and robust evidence of a positive and significant effect of diversity among college-educated immigrants on GDP per capita. Conversly, we find no impact of low-skilled diversity or contamination effects.

Chapter 3 fits within the literature looking at the determinants of the performance of immigrants in the destination country labor markets. We investigate how natives' attitudes affect immigrants' unemployment duration in Germany. Using individual level panel data from the German Socio Economic Panel from 1984 to 2012, we use survival analysis methods to model immigrants' unemployment durations. We find that lower trust levels of natives towards the citizens of a given country, measured using Eurobarometer surveys, positively influence the unemployment duration of immigrants originating from this country. Our results highlight the fact that immigrants face different obstacles depending on their origin when it comes to integrating destination country labor markets.

Chapter 4 analyses whether international migrants contribute to increasing technological advances in developing countries by inducing a transfer of productive knowledge from developed countries back to migrants' home countries.

Using the Economic Complexity Index as a proxy for the amount of productive knowledge embedded in each countries and bilateral migrant stocks of 20 OECD destination countries, we show that international migration is a strong channel of technological transmission.

Keywords: Economic Development, International Migration, Immigrant Workers, Birthplace Diversity, Growth, Unemployment Duration, Discrimination, Technological Change.

JEL codes: F63, F22, J61, J64, J71, 033.

List of acronyms

ACS:	American Community Survey
AFT:	Accelerated Failure Time
CERDI:	Centre d'Études et de Recherches sur le Développement International
CZs:	Communting Zones
Dep:	Dependent Variable
ECI:	Economic Complexity Index
ESS:	European Social Survey
FE:	Fixed Effects
FDI:	Foreign Direct Investment
GDP:	Gross Domestic Product
GMM:	Generalized Method of Moments
GSOEP:	German Socio-Economic Panel
IAB:	Institut flür Arbeitsmarktund Berufsforchung
IPUMS:	Integrated Public Use Microdata Series
IV:	Instrumental Variable
NHL:	North American National Hockey League
OECD:	Organization for Economic Cooperation and Development
OLS:	Ordinary Least Squares
PPML:	Pseudo Poisson Maximum Likelihood
UNHCR:	United Nations High Commissioner for Refugees
UNRWA:	United Nations Relief and Work Agency
RCA:	Revealed Comparative Advantage
SITC:	Standard International Trade Classification
US:	United States
WDI:	World Development Indicators
WVS:	World Value Surveys
WW2:	World War II
2SRI:	Two-Stage Residual Inclusion

Contents

List of Figures

List of Tables

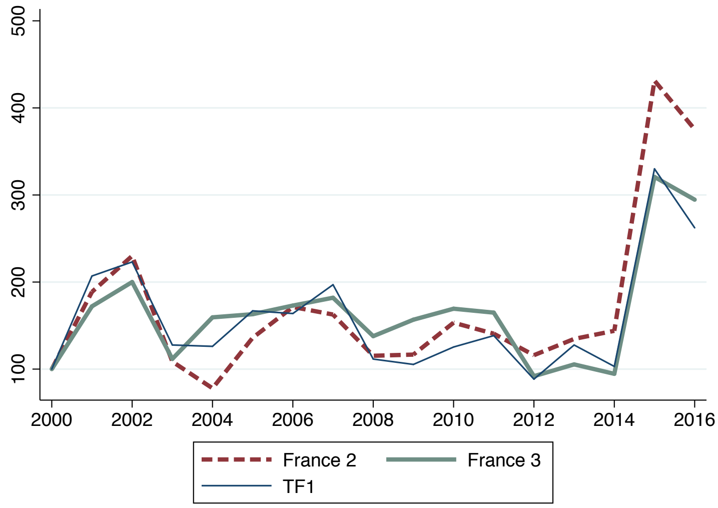
Chapitre 1
Introduction générale

"Immigrants are not just labor inputs [...] immigrants are people."
George J. Borjas, *We Wanted Workers*.

H. est né au Burkina Faso dans la petite province de Namentenga. Après une enfance passée dans son pays d'origine, H. a décidé de tout quitter à 19 ans pour immigrer en France, afin de terminer ses études d'économie. Aujourd'hui, cela fait 6 ans que H. vit en Europe. Durant ces quelques années H. a travaillé, consommé et participé à la vie sociale de son pays d'accueil, à travers notamment son engagement dans différentes associations locales. Durant ce séjour, il a interagi avec les natifs de son pays d'accueil, nouant des liens d'amitié profonds d'une part, et faisant aussi parfois face à des attitudes hostiles à sa venue d'autre part. Pendant tout ce temps, H. n'a pas rompu les liens avec son pays d'origine. En 6 ans, il a envoyé plus de 6000 dollars à sa famille et compte bien un jour revenir dans son pays pour mettre à profit ses talents et son expérience acquise en Europe. L'histoire de cette thèse c'est l'histoire de H. mais aussi de millions de migrants à travers le monde. C'est l'histoire d'étrangers, arrivés à destination avec leur éducation, leur culture et leurs normes, et de la diversité qu'ils apportent à leur pays d'accueil. C'est aussi l'histoire de leurs échanges et de leurs interactions avec les natifs des pays dans lesquels ils ont choisi de venir s'installer. C'est enfin l'histoire d'hommes et de femmes qui gardent un attachement fort avec leur pays d'origine et servent de pont entre les nations. C'est l'histoire de mouvements qui sont parfois anciens et qui, pourtant, ne cessent toujours de nous questionner sur leurs implications économiques et sociales.

Les travaux présentés dans cette thèse arrivent à un moment où l'attention pour le sujet des migrations n'a jamais été aussi grande. La Figure 1.1 fait parfaitement état de cela en présentant le nombre de sujets diffusés dans les

FIGURE 1.1: Évolution de la couverture médiatique des journaux télévisés Français sur l'immigration



Source : Calculs de l'auteur à partir des données de l'institut national de l'audiovisuel Français (INA). Nombre annuel de sujets traités dans les principaux journaux télévisés Français, base 100 en 2000, comprenant un des mots clés suivants : migrant(s), migration, immigration et/ou réfugié(s). Le nombre de sujets en 2000 est respectivement de 130, 150 et 111 pour « le journal de 20 heures » de TF1, « le journal de 20 heures » de France 2 et le « 19/20 » de France 3.

principaux journaux télévisés Français comprenant au moins un des mots suivants : « migrants, migration, immigration ou encore réfugiés ». L'année 2015, tout particulièrement, marque un tournant sans précédent dans le traitement médiatique du sujet des migrations, en raison de l'augmentation récente des flux migratoires en Europe due à la multiplication des conflits en Afrique et au Moyen-orient notamment. Pour cette seule année, plus de 1,2 millions de réfugiés et de migrants sont arrivés en Europe[1].

Cette introduction générale vient brièvement présenter le sujet des migrations internationales et synthétiser les avancées de la recherche économique sur les questions dont traite cette thèse. Il ne convient évidemment en aucun cas de fournir une analyse exhaustive de la littérature mais plutôt d'insérer les contributions des différents Chapitres de cette thèse dans la littérature économique sur les migrations internationales. Il est important de noter que ces trois

[1]Chiffres de l'Agence des Nations Unies pour les réfugiés (UNHCR) pour l'année 2015.

chapitres sont de nature empirique bien qu'ils empruntent tous trois leurs intuitions de la théorie économique. Les approches macro et micro économiques sont alternativement utilisées tant elles apparaissent complémentaires pour le sujet des migrations internationales. Nous utilisons ainsi à la fois des données individuelles d'enquêtes ou de recensements ainsi que des données agrégées lorsque cela est nécessaire. Une grande place est aussi accordée aux stratégies d'identification. Les implications des travaux menés dans cette thèse, ainsi que les recommandations de politiques économiques associées, sont évoquées en conclusion.

L'introduction ci-dessous s'articule de la manière suivante : La section 1.1 dresse, après avoir rappelé quelques notions essentielles, un portrait de l'état des migrations internationales dans le monde pour l'année 2015. Le constat de grandes disparités géographiques dans la répartition des migrants, nous amène alors à repenser l'évolution des migrations internationales sur les 50 dernières années, et ce, avec un regard tout particulier sur les pays de l'OCDE dans la section 1.2. Il ressort de cette analyse statistique préliminaire que, non seulement les pays de l'OCDE font face à une augmentation de la part des immigrés dans leur population depuis 1960, mais aussi que cette augmentation est majoritairement tirée par l'arrivée de migrants en provenance de pays en développement. L'augmentation de la diversité culturelle induite par ce double phénomène est ainsi au cœur de la section 1.3. La section 1.4 présente, quand à elle, quelques faits stylisés sur les perceptions que peuvent avoir les natifs des pays de l'OCDE envers l'immigration, et leurs conséquences. En effet, il est probable que les attitudes des natifs occupent une part importante dans le processus d'assimilation économique des migrants, conditionnant l'impact des ces derniers sur leur pays d'accueil. Enfin, l'intégration des migrants dans leur pays d'accueil est aussi à l'origine des retombées attendues dans leur pays d'origine. Ces retombées font l'objet de la section 1.5 qui vient conclure l'introduction de cette thèse.

1.1 Un bref état des lieux

De la définition des migrants au niveau international

Employée à tord et à travers, dans les médias notamment, la notion de migrant international n'est pas toujours évidente et il nous convient donc ici de la définir. Les Nations Unies définissent un migrant international comme *« toute personne qui change son pays de résidence usuel »*. Tout individu répondant à ce critère est alors considéré comme un immigré dans son pays d'accueil et comme un émigré pour son pays d'origine. Cette première définition, par l'emploi du terme « usuel », exclut donc les voyages d'affaires ou personnels, tout comme les déplacements religieux ou à raison médicale, de la notion de migrations internationales. D'autres nuances sont ajoutées par les Nations Unies comme la distinction entre migrations de long et court terme. En effet, sont considérées comme migrant de long terme, toutes les personnes ayant changé de lieu usuel de résidence pour au moins 12 mois. Tout changement de lieu de résidence supérieur à trois mois et inférieur à 12 mois, à l'exception des déplacements temporaires évoqués précédemment, est alors considéré comme une migration de court terme.

Il est intéressant de noter que la définition des migrants internationaux par les Nations Unies ne fait aucunement référence au motif de la migration. Cependant, les traités internationaux introduisent une séparation nette entre la définition d'un migrant et d'un réfugié. En effet, le statut de réfugié est encadré par la convention des Nations Unies relatif au statut des réfugiés, signé à Genève en 1951, qui définit comme réfugié *« toute personne craignant avec raison d'être persécutée du fait de sa race, de sa religion, de sa nationalité, de son appartenance à un certain groupe social ou de ses opinions politiques, se trouve hors du pays dont elle a et qui ne peut ou, du fait de cette crainte, ne veut se réclamer de la protection de ce pays ; ou qui, si elle n'a pas de nationalité et se trouve hors du pays dans lequel elle avait sa résidence habituelle à la suite de tels évènements, ne peut ou, en raison de ladite crainte, ne veut y retourner »*. Contrairement aux migrants internationaux, le principe de non refoulement interdit les états signataires d'expulser un réfugié dans son pays d'origine. Pour l'année 2015, le Haut Commissariat des Nations Unies pour les réfugiés reporte 16,1 millions de réfugiés dans le monde (23,1 millions en comptant les réfugiés Palestiniens sous mandant du UNRWA)[2]. Il convient d'ores et déjà de préciser que, même si de nombreux résultats de cette thèse, et leurs implications, conviennent à la fois aux migrants et aux réfugiés, les travaux ici réalisés ne se concentrent pas spécifiquement sur la question des réfugiés mais sur les migrations au niveau international. Nous ne nous intéressons pas non plus aux migrations internes qui, contrairement aux migrations internationales, n'impliquent pas la traversée de frontières nationales. Il est cependant intéressant de noter que le Haut Commissariat des Nations Unies pour les réfugiés (UNHCR) rapporte un record annuel de 38 millions de personnes déplacées, au sein de leur propre pays, en 2015. Cela correspond à une augmentation de 470% par rapport à l'année 2005[3].

De la définition théorique aux statistiques

Si la définition des Nations Unies est un point de départ, elle ne permet pas en l'état l'analyse statistique des migrations internationales. En effet, établir un recensement des migrants dans le monde requiert des précisions supplémentaires qu'il convient ici d'aborder. La plupart des données internationales sur la migration sont collectées à destination dans les pays de l'OCDE. Ces données sont issues des registres de populations (registres dans lesquels les étrangers doivent obligatoirement s'inscrire à leur arrivée), des recensements, des permis de séjour délivrés ou encore d'enquêtes. Le principal problème posé par la définition des Nations Unies est son caractère abstrait et le nombre d'informations différentes qu'elle nécessiterait pour être effectivement utilisée. Deux proxies sont alors employées pour définir le groupe des migrants dans un pays donné.

[2]La répartition des réfugiés dans le monde est assez inégale puisque 53% des réfugiés sont originaires de seulement 3 pays à savoir la Somalie (1,1 millions), l'Afghanistan (2,7 millions) et la Syrie (4,9 millions). Les principaux pays récipiendaires de réfugiés sont la Turquie (2,5 millions), le Pakistan (1,6 millions), le Liban (1,1 millions) ou encore la République islamique d'Iran (1 million). La moitié des réfugiés dans le monde ont moins de 18 ans.

[3]Une confusion supplémentaire est souvent introduite par l'utilisation du terme réfugié climatique. En effet ce terme, qui n'a pas de fondements juridiques, ne répond en aucun cas à la définition des réfugiés donnée par les Nations-Unies. Le 20 juillet 2015, la Cour suprême de Nouvelle-Zélande a ainsi refusé le statut de réfugié climatique à un habitant des îles Kiribati ayant fui la montée des eaux dans son pays d'origine. À l'heure actuelle, aucune statistique officielle n'enregistre clairement les migrants climatiques. Les estimations disponibles varient entre 25 millions et 1 milliard de migrations climatiques attendues pour l'année 2050.

Majoritairement, c'est le pays de naissance qui est retenu. C'est la définition qui est par exemple utilisée en France, aux États-Unis ou encore en Espagne. En revanche, d'autres pays, comme l'Allemagne, l'Italie ou le Japon, définissent un migrant sur la base de la nationalité. Cette deuxième définition apparait moins judicieuse en raison de son caractère changeant. En effet, alors que le pays de naissance ne varie pas dans le temps, la nationalité peut-être obtenue à destination par exemple. L'emploi de définitions différentes pose évidemment le problème de la comparaison entre pays, même si ce dernier n'est que rarement soulevé[4]. L'évolution des frontières au cours du temps, les valeurs manquantes, la désagrégation des données par niveaux d'éducation et par age ou encore, le fait que la plupart des recensements n'identifient pas clairement le pays d'origine des migrants mais une région plus large, sont autant de challenges que les bases de données sur les migrations internationales doivent relever (Docquier et Marfouk, 2006; Özden *et al.*, 2011; United Nations, 2015) et qui rendent encore les données sur les migrations internationales perfectibles.

De nombreuses améliorations sont alors encore souhaitables dans la mesure de la mobilité humaine au niveau international. La plupart des pays en développement n'ont en effet que peu, ou pas, de données fiables pour la migration internationale quand, pour les pays développés, les accords de libre circulation rendent toute mesure difficile. De plus, les migrants illégaux échappent la plupart du temps à l'enregistrement des flux internationaux de migrants. Enfin, alors que les chercheurs seraient intéressés par les flux de migrants, qui font sens pour la majeure partie des modèles théoriques existants en économie des migrations, nous ne disposons le plus souvent que des stocks de migrants observés. Or, les variations dans les stocks de migrants ne peuvent qu'imparfaitement refléter les flux de migrants, puisque ne prenant pas en compte, non seulement les migrations retours, mais aussi les évènements démographiques comme les décès par exemple (Docquier et Rapoport, 2012). L'émergence des données en provenance du Web devrait pouvoir, à l'avenir, résoudre ces questions (Messias *et al.*, 2016).

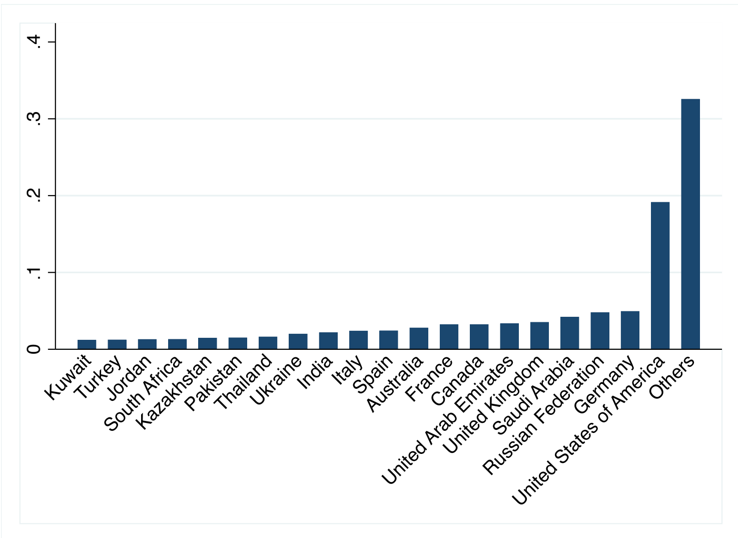
[4]Les deux bases de données utilisées dans cette introduction (Özden *et al.*, 2011; United Nations, 2015) utilisent, lorsque cela est possible, le pays de naissance pour définir la migration. Pour les pays où seule la nationalité est disponible, c'est cette dernière définition qui est retenue.

De l'état des migrations internationales en 2015 et leur importance

Combien dénombre-t-on alors de migrants dans le monde, sur la base des mesures définies précédemment ? Selon la Division Population des Nations Unies, le nombre de migrants en 2015 s'établissait à 244 millions (United Nations, 2015). Si à première vue ce chiffre peut paraitre important il devient négligeable lorsqu'il est mis en perspective avec la population mondiale. En effet, le nombre de 244 millions de migrants apparait faible lorsque l'on sait que cela correspond à seulement 3,3% de la population mondiale. De plus, penser le chiffre des migrations internationales n'a de sens que si ce dernier est mis en comparaison avec les écarts de revenus observés au niveau mondial. En effet, sachant que les différences de revenus entre pays sont un des principaux déterminants de la mobilité, et que, moins d'un habitant de la planète sur deux vit avec moins de 2 dollars par jour, il apparait évident que des freins importants entravent la mobilité des individus au niveau international. McKenzie *et al.* (2010), en utilisant les résultats d'une expérience naturelle sur la migration entre le Royaume des Tonga et la Nouvelle-Zélande, estiment à 263% les gains dus à la mobilité entre ces deux pays, et ce seulement un an après la migration. Pour les États-Unis, Clemens *et al.* (2010) montrent que, en moyenne, les ressortissants de 42 pays en développement obtiennent une prime équivalente à 10000 dollars par rapport au salaire qu'ils auraient obtenu s'ils étaient restés dans leur pays d'origine. Ce dernier chiffre correspond à un doublement du revenu par tête des pays en développement. Dans ce contexte, il n'est pas surprenant alors que Pritchett (2009) évoque l'image de « précipice aux frontières » ("cliff at the border"). Bertoli et Fernandez-Huertas Moraga (2015), qui mesurent la taille de ces précipices, montrent que la mise en place d'un visa d'un pays donné sur un tiers entraine une diminution du niveau des flux migratoires entre 40 et 47% en provenance de ce dernier. Aux politiques migratoires restrictives viennent bien sûr s'ajouter les coûts directs de la migration ainsi que les contraintes informationnelles et de crédits qui réduisent de manière drastique la mobilité internationale (Hatton et Williamson, 2006; Angelucci, 2015).

Face à de tels gains, et, compte tenu du faible nombre de migrants au niveau international, il est naturel de se demander quels seraient les flux de migrants si les entraves à la migration étaient complètement levées. La réponse à cette question se trouve partiellement dans les enquêtes « Gallup World Poll » menées dans 148 pays entre 2007 et 2010. Ces enquêtes reportent le nombre d'individus qui désireraient migrer s'ils le pouvaient. Les résultats montrent qu'environ 526 millions de personnes migreraient s'ils en avaient la possibilité (12% de la population mondiale). Pour ces migrants potentiels, les États-Unis apparaissent comme la destination la plus attractive (24%) devant le Canada (7%), le Royaume-Uni (7%) ou encore la France (6%) (Docquier *et al.*, 2015). D'une manière générale, seulement 15 pays attireraient 70% des flux reportés. Là encore, la faiblesse de ces chiffres peut paraitre surprenante mais s'explique assez facilement par les coûts indirects de la migration comme les coûts psychologiques par exemple (Borjas, 2015).

FIGURE 1.2: Répartition des migrants dans le monde en 2015 en pourcentage de la migration totale

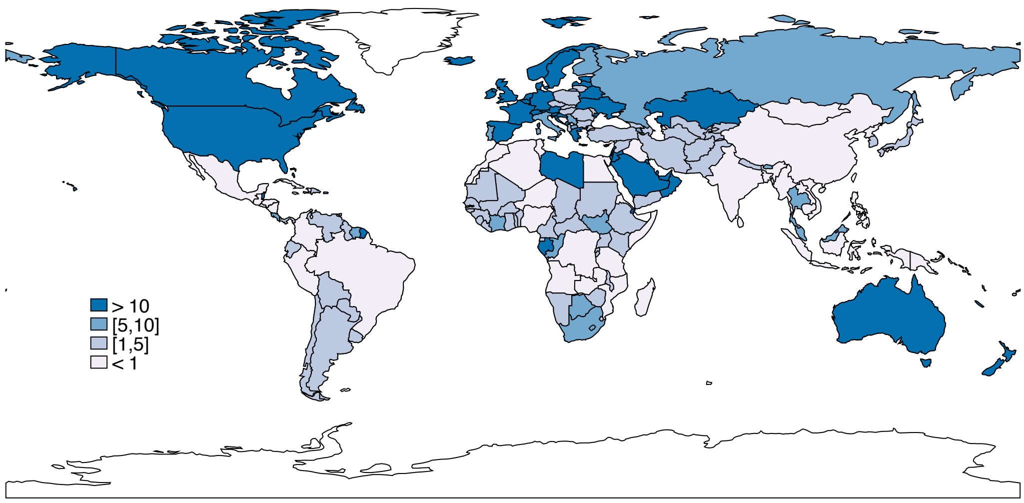


Source : Calculs de l'auteur à partir des données de la division population des Nations Unies "Trends in international migrant stock : The 2015 revision". Seuls les pays accueillant 2/3 du total des migrants à travers le monde sont représentés sur le graphique. Les autres pays sont rassemblés sous la mention « others ».

Une répartition inégale

Derrière ces quelques chiffres sur l'état des migrations internationales au niveau mondial se cachent évidemment de nombreuses disparités. Comme reporté dans le Figure 1.2, 20% des migrants internationaux résident aux États-Unis quand les 2/3 du stocks international de migrants se concentre exclusivement dans 20 pays receveurs. Cette inégale répartition se traduit par des différences majeures dans les taux d'immigration observés au niveau mondial. En effet, alors que les pays développés d'Amérique du Nord, d'Europe ou encore d'Océanie ont des taux d'immigration en moyenne supérieurs à 10%, les pays d'Afrique, d'Asie ou encore d'Amérique Latine présentent en moyenne des taux d'immigration inférieurs à 2% de leur population totale (Figure 1.3). Ainsi, la part de la population nait à l'étranger atteint 12,1% en France, 13,2% au Royaume Uni, 14,5% aux États-Unis ou encore 15% en Allemagne. Au niveau international, le taux d'immigration le plus faible va à la Chine avec moins de 0,1% de migrants alors que le record est pour les Émirats Arabes Unis pour lesquels la part de la population nait dans un pays étranger atteint 88%.

FIGURE 1.3: Taux d'immigration dans le monde en 2015.



Source : Calculs de l'auteur à partir des données de la division population des Nations Unies "Trends in international migrant stock : The 2015 revision". Le taux d'immigration est calculé comme le ratio entre le nombre total de personnes nées à l'étranger et la population totale du pays de destination.

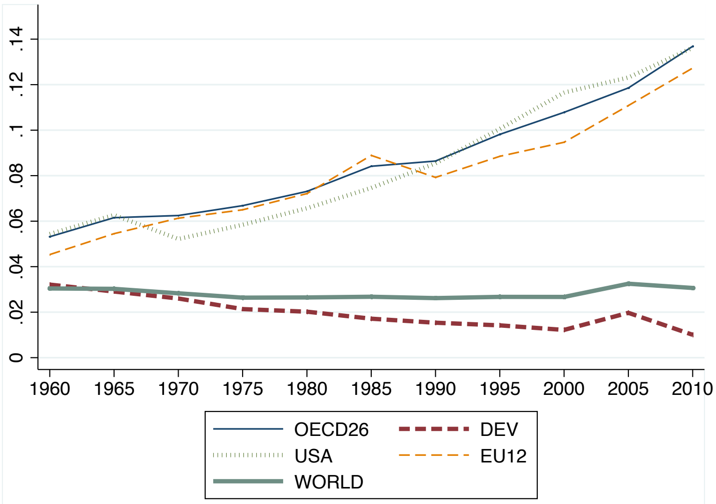
En conclusion, si une analyse statique souligne le faible nombre de migrants au niveau international, un zoom rapide sur les pays de l'OCDE met en évidence de fortes disparités dans leur répartition. Les pays de l'OCDE semblant les plus touchés par le phénomène d'immigration, il convient alors de regarder l'évolution et les tendances des migrations internationales sur les 50 dernières années dans ces pays afin de comprendre comment le sujet des migrations a vu son importance grandir jusqu'à devenir un sujet central non seulement de la discussion économique mais aussi de l'actualité.

1.2 Évolution des migrations internationales sur 50 ans

L'analyse du stock international de migrants sur les 50 dernières années, sans distinction sur les pays d'acceuil, ne révèle aucun changement majeur dans la mobilité entre pays. En effet, si, entre 1960 et 2015, le stock international de migrants a bien progressé, passant de 91.6 à 244 millions, ce dernier, rapporté à la population mondiale, est resté stable, fluctuant autour de 3%[5]. Entre 1960 et 2015, la part de la migration internationale est passée de 3,05 à 3,32% ce qui correspond ni plus ni moins à une augmentation de environ 1 point de pourcentage par rapport aux niveaux observés au début du 20$^{\text{ème}}$ siècle (McKeown, 2004). Comme énoncé précédemment, ces chiffres cachent néanmoins de nombreuses disparités entre régions et groupes de revenus. Comme le montre la Figure 1.4, la part des migrants dans la population totale des pays en développement a légèrement diminué passant de 2,3% à 1.1%. En revanche, la même analyse pour les pays à haut revenu montre une augmentation importante de 4,9 à 11.7% (de 3.9 à 12.2% pour l'Union Européenne des 15, de 5.4% à 13,6% pour les États-Unis, ou encore de 15 à 22% pour l'Australie ou le Canada). De plus, l'augmentation de la part des migrants dans la population des pays de l'OCDE a été entièrement tirée par la migration Sud-Nord, c'est à dire en provenance des pays en développement (Figure 1.5). La part des individus nés dans un pays en développement et résidant dans un pays de l'OCDE est passée

[5]Il est important de noter que ces chiffres sont ajustés pour l'effondrement du bloc soviétique et ne prennent pas systématiquement en compte les demandeurs d'asiles.

FIGURE 1.4: Taux d'immigration, 1960-2010 (en pourcentage de la population totale)



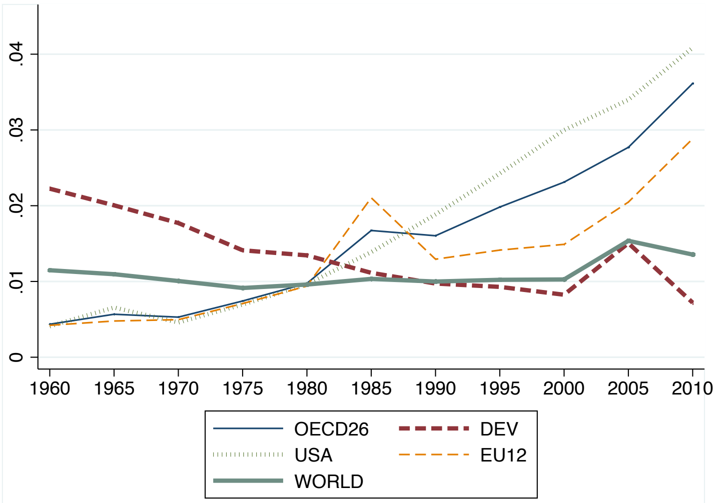
Source : Calculs de l'auteur à partir de Özden *et al.* (2011) et United Nations (2015). Le taux d'immigration est calculé comme le ratio entre le nombre total de personnes nées à l'étranger et la population totale du pays de destination. Les années 1965, 1975, 1985, 1995 et 2005 sont estimées par interpolation linéaire.

de 2,0 à 8,7% en l'espace d'un demi-siècle. En moyenne, les migrants originaires de pays en développement comptent pour 98% de l'augmentation de la migration de la population des pays à haut revenu sur la période 1960-2010 (80% en Europe, 120% aux États-Unis ou encore 150% en Australie et au Canada)[6].

En conclusion, il semble donc que deux faits majeurs émergent de l'analyse descriptive des chiffres sur la migration internationale en longue période à savoir (i) une augmentation de la part des migrants dans la population des pays de l'OCDE depuis 1960 et (ii) une augmentation de l'immigration en provenance de pays géographiquement, économiquement et culturellement de

[6]Un deuxième fait stylisé observable, à propos de la migration Sud-Nord des 50 dernières années, est l'augmentation de la part des diplômés du tertiaire dans la migration totale. Docquier et Rapoport (2012) documentent cette tendance et ses effets sur les pays d'origine des migrants. Ils montrent notamment que, sous certaines conditions, les effets négatifs du départ des plus éduqués des pays en développement vers les pays de l'OCDE, peuvent être compensés par une augmentation du niveau d'éducation globale dans les pays d'origine. En effet les retombées économiques de la migration et les plus grandes facilités à migrer peuvent pousser les individus des pays en développement à investir de manière plus forte dans l'éducation, sans que toutefois, à la fin, tous aient la possibilité de quitter leur pays.

Figure 1.5: Taux d'immigration en provenance de pays en développement, 1960-2010 (en pourcentage de la population totale)



Source : Calculs de l'auteur à partir de Özden *et al.* (2011) et United Nations (2015). Le taux d'immigration en provenance des pays en développement est calculé comme le ratio entre le nombre total de personnes nées à l'étranger et la population totale du pays de destination. La liste des pays en développement est issue de la classification de la Banque Mondiale en 2015. Les années 1965, 1975, 1985, 1995 et 2005 sont estimées par interpolation linéaire.

plus en plus distants de ces derniers.

1.3 De l'impact des migrants sur leur pays d'accueil

Une littérature controversée

Le constat d'une immigration croissante dans les pays de l'OCDE, en provenance de pays en développement entre 1960 et 2010, a soulevé le débat, à la fois dans la littérature économique, et dans la société civile, sur son effet dans les pays d'accueil. Résumer une littérature aussi vaste que celle de l'impact de l'immigration dans les pays d'accueil remplirait un objectif qui irait bien au delà de celui de cette thèse, tant cette littérature est vaste et soulève encore de nombreuses questions. La liste des travaux cités ci-dessous ne représente alors en aucun cas une sélection exhaustive. L'exercice est d'autant plus difficile que l'immigration peut avoir des impacts sur la fiscalité[7], le bien-être des natifs, la productivité ou encore la cohésion sociale pour ne citer qu'eux ; autant de canaux différents aux travers desquels l'immigration peut affecter l'économie. Alors que l'idée de milliards de dollars perdus, en raison des restrictions à la mobilité, revient régulièrement dans la littérature (Clemens, 2011), d'aucuns sont plus méfiants quant à la libéralisation des flux humains en raison de l'entremêlement des effets économiques et sociaux, souvent mal connus de l'immigration (Borjas, 2015).

Le canal le plus traité dans la littérature sur les effets économiques de l'immigration est assurément l'impact des migrants sur le marché du travail et particulièrement sur les natifs (Card, 1990; Friedberg et Hunt, 2015; Borjas *et al.*, 2012; Card, 2001; Borjas, 2003; Cortes, 2008; Ottaviano et Peri, 2012; Borjas *et al.*, 2012). Deux grands résultats s'opposent sur les effets que peut avoir l'immigration sur le marché du travail. D'un côté, des auteurs comme Borjas (2003) montrent un impact négatif de l'immigration sur les salaires des natifs aux États-Unis, en concurrence directe avec des migrants ayant des mêmes niveaux d'éducation et d'expérience, pour la période 1960-2000. Une augmentation de 10% de l'offre au sein d'un groupe particulier de travailleurs entraine une réduction des salaires de 3 à 4% en moyenne et réduit l'offre de travail des natifs (nombre annuel de semaines travaillées divisé par 52) de environ 3,7 points de pourcentage. A l'inverse, Ottaviano et Peri (2012) soulignent l'importance de prendre en compte l'élasticité de substitution entre migrants et natifs et la réponse des investissements en capital physique aux

[7]Une question au cœur de l'analyse économique de l'impact de l'immigration est son influence sur les finances publiques des pays d'accueil. La contribution des migrants au système fiscal dépend de la structure de l' immigration en terme d'âge, d'éducation et de l'assimilation économique des migrants. Barbone et al. (2009), pour 13 pays de l'Union Européenne, mettent en évidence une contribution positive nette des migrants au système fiscal, excepté pour les ménages où les immigrés sont originaires d'autres pays Européens. La contribution moyenne des ménages avec un migrant est d'environ 1000 euros par an et par personne. Une étude plus récente montre en revanche un impact plutôt négatif de l'immigration sur la fiscalité des pays receveurs de l'OCDE (OECD, 2013). Les migrants semblent contribuer de manière plus faible au système social et tendent à en être un peu plus dépendants que les natifs. Or, au niveau agrégé, cet impact est très faible et sans conséquence sur la contrainte budgétaire des États. L'effet net se situe en effet entre - 2 et + 2% du PIB selon les estimations.

changements dans la structure des qualifications de la population dans une approche d'équilibre général. Ces deux auteurs estiment alors un effet global positif de l'immigration, aux États-Unis entre 1990 et 2004, pour l'ensemble des groupes de travailleurs, hormis les travailleurs déscolarisés avant la fin du lycée pour lesquels l'impact est négatif. En moyenne l'effet sur le salaire des natifs est de +0.6% pour une augmentation de l'immigration similaire à celle de Borjas (2003). À elles seules, ces deux études montrent bien la difficulté pour les économistes de parvenir à un consensus sur l'effet global de l'immigration sur les salaires des natifs[8]. En revanche, il semble que cet effet, qu'il soit négatif ou positif, et ce, de manière unanime, soit de faible ampleur.

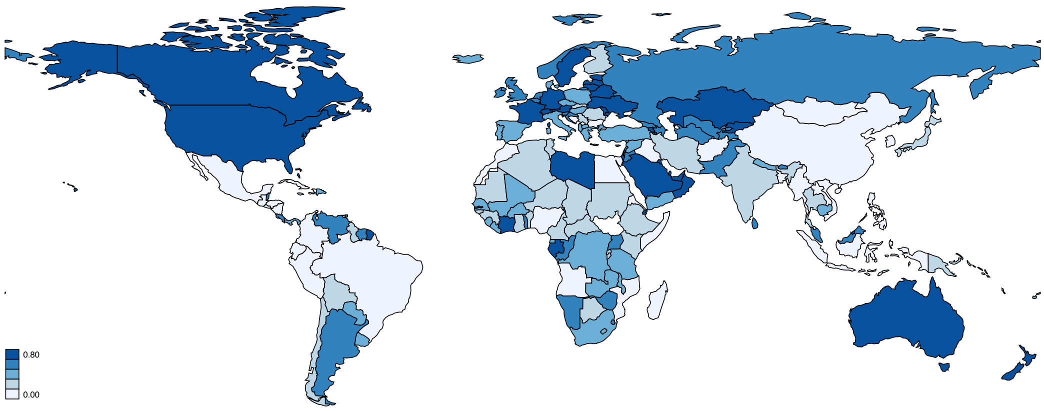
De l'effet taille de l'immigration à la diversité

Face à l'impossibilité de conclure quant à un effet clair de l'immigration sur l'économie, une partie des économistes ont alors déplacé leur regard non plus vers un effet « taille » de l'immigration mais vers un effet « variété[9] ». En effet, l'augmentation continue du stock de migrants venant de pays en développement a mécaniquement entrainé une augmentation de la diversité culturelle dans les pays à haut revenu. Or, mesurer la diversité culturelle est évidemment difficile de par son inhérente immatérialité. Pour ce faire, les économistes ont alors recours au pays de naissance des individus. L'hypothèse centrale de cette littérature est donc que la culture des individus, leurs normes, leur éducation ou encore leurs compétences sont majoritairement déterminées par leur pays de naissance.

[8] Un sous-groupe de papiers dans cette littérature utilisent des expériences naturelles pour estimer l'impact de l'arrivée de migrants sur un marché du travail donné (Card, 1990; Hunt, 1992; Carrington et Lima, 1996; Friedberg, 2001; Clemens, 2013; Borjas, 2017). Card (1990) estime l'impact de l'arrivée de migrants Cubains sur le marché du travail de Miami, lors de l'épisode de l'exode de Mariel d'avril 1980, où, en moins de 3 mois, plus de 100 000 Cubains débarquèrent par bateaux dans la ville (augmentation de plus de 7% de la force de travail). En comparant la ville de Miami avec d'autres villes des USA non impactées par ce choc, Card (1990) ne trouve aucun effet de l'immigration sur les salaires des natifs, ni d'effets sur les salaires des migrants Cubains déjà installés en ville. Borjas (2017) pour le même épisode, estime un effet négatif de l'immigration sur le salaire des natifs et une élasticité des salaires au nombre de travailleurs de -0.5 à -1.5.

[9] Si l'emploi du terme « variété » pour parler de l'origine des migrants peu paraitre inapproprié, celui-ci est volontairement utilisé ici pour mettre en parallèle la littérature sur la migration avec celle sur le commerce international.

FIGURE 1.6: Diversité moyenne parmi les résidents selon le pays de naissance, 1960-2010

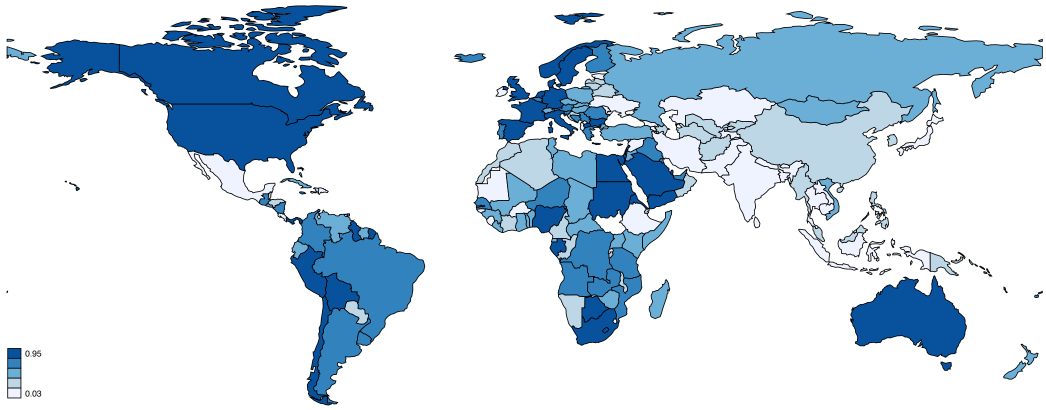


Notes : Diversité moyenne parmi les résidents entre 1960 et 2010. Elle est définie comme la probabilité deux tirer aléatoirement, dans la population totale, deux individus aux pays de naissance différents. Source : Calculs de l’auteur à partir de Özden *et al.* (2011) et United Nations (2015).

La Figure 1.6 trace les contours de la diversité dans les différents pays du monde en 2010. La diversité est ici définie comme la probabilité de tirer aléatoirement deux individus dans la population des résidents et que ces deux individus soient nés dans deux pays différents. Les pays de l’OCDE sont alors, sur la base du pays de naissance de leur résident, des pays plus fortement diversifiés que les autres. Or, et comme énoncé précédemment, ces pays comptent en moyenne 10% de migrants et la diversité, telle que définie ici, revient donc à mesurer la probabilité de tirer un natif et un immigré, quelque soit le pays de naissance de ce dernier. Cette mesure est donc fortement corrélée avec le taux d’immigration et ne permet de distinguer l’effet « taille » de l’effet « variété » de l’immigration. Comparer visuellement les Figures 1.3 et 1.6 permet de vérifier aisément cela. Alesina *et al.* (2016) proposent donc de définir une mesure de la diversité à l’intérieur du groupe des migrants (Figure 1.7). Tout en contrôlant pour le taux d’immigration, cette mesure permet alors de mieux appréhender les effets de la diversité sur les pays d’accueil des migrants.

De manière purement théorique, la diversité peut avoir des effets positifs

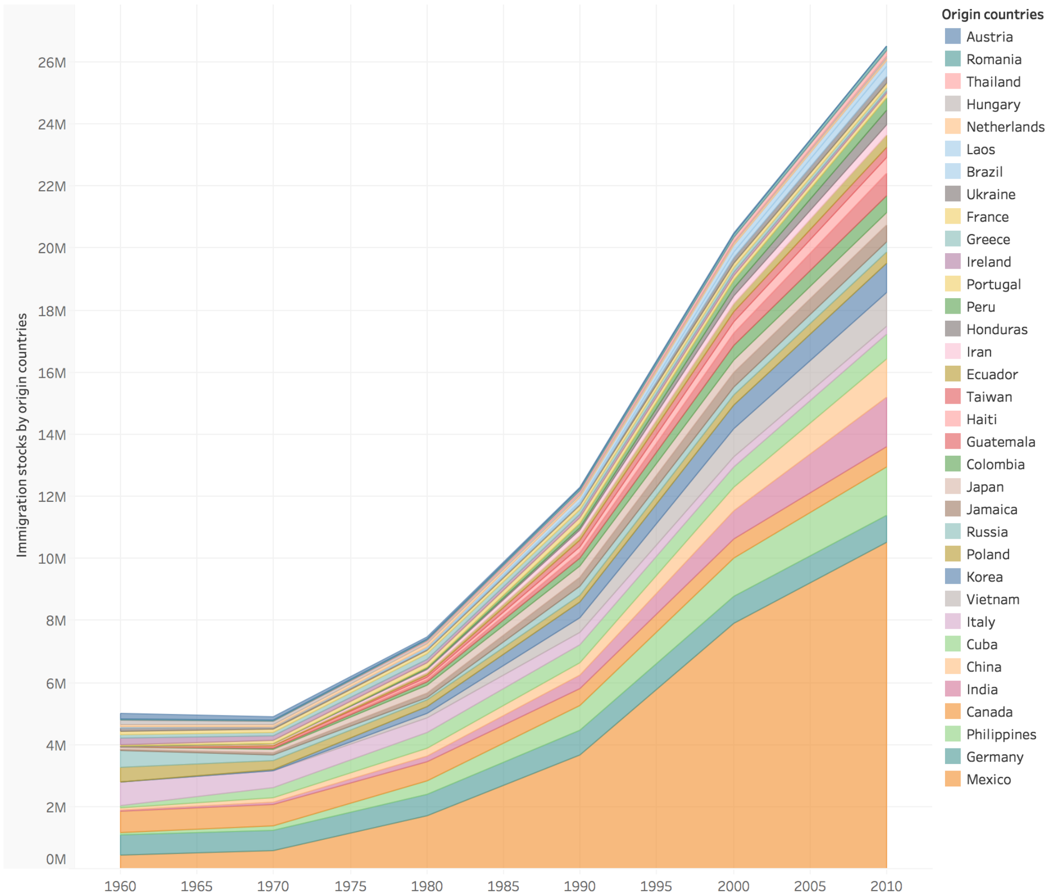
Figure 1.7: Diversité moyenne parmi les immigrés selon le pays de naissance, 1960-2010



Notes : Diversité moyenne parmi les immigrés entre 1960 et 2010. Elle est définie comme la probabilité deux tirer aléatoirement, dans le groupe des migrants, deux individus aux pays de naissance différents. Source : Calculs de l'auteur à partir de Özden *et al.* (2011) et United Nations (2015).

au travers de l'augmentation de la productivité des firmes. Celles-ci peuvent, en effet, bénéficier au travers d'une plus grande diversité culturelle, d'idées originales et innovantes facilitant la résolution de problèmes, d'une meilleure connaissance du marché mondial ou encore d'une complémentarité accrue entre travailleurs natifs et migrants. Au niveau macro-économique, la diversité peut résoudre les problèmes de sous capacité d'emploi dans certains secteurs ou encore créer de nouvelles complémentarités entre différents secteurs. Évidemment, accueillir de plus en plus de personnes d'origines diverses peut aussi entraver la communication, détériorer la confiance entre les individus et détruire les liens sociaux et donc, être un frein à la croissance. L'entremêlement de ces effets à la fois positifs et négatifs fait donc de cette question un problème par définition empirique nécessitant de dégager un effet net de la diversité sur l'économie. Différents auteurs se sont appliqués à estimer l'impact de la diversité culturelle sur l'économie (Ottaviano et Peri, 2006; Boeheim *et al.*, 2012; Ager et Brückner, 2013; Kahane *et al.*, 2013; Ozgen *et al.*, 2014; Parrotta *et al.*, 2014; Suedekum *et al.*, 2014; Fulford *et al.*, 2015; Alesina *et al.*, 2016). Leurs contributions respectives sont plus largement détaillées dans le Chapitre 2 de cette thèse.

FIGURE 1.8: Immigration aux États-Unis entre 1960 et 2010



Notes : Stocks d'immigrés pour les pays représentant, en moyenne sur la période, plus de 0.5% de la migration totale. Source : Calculs de l'auteur à partir des données IPUMS.

Contribution de la thèse : De l'effet de la diversité en longue période

Le Chapitre 2 de cette thèse contribue à cette littérature en estimant, pour la première fois, l'effet de la diversité culturelle à l'intérieur du groupe des migrants, avec des données[10] de panel en longue période. En effet, notre étude

[10] Les données de cette étude proviennent des données de recensement Américain (Integra-

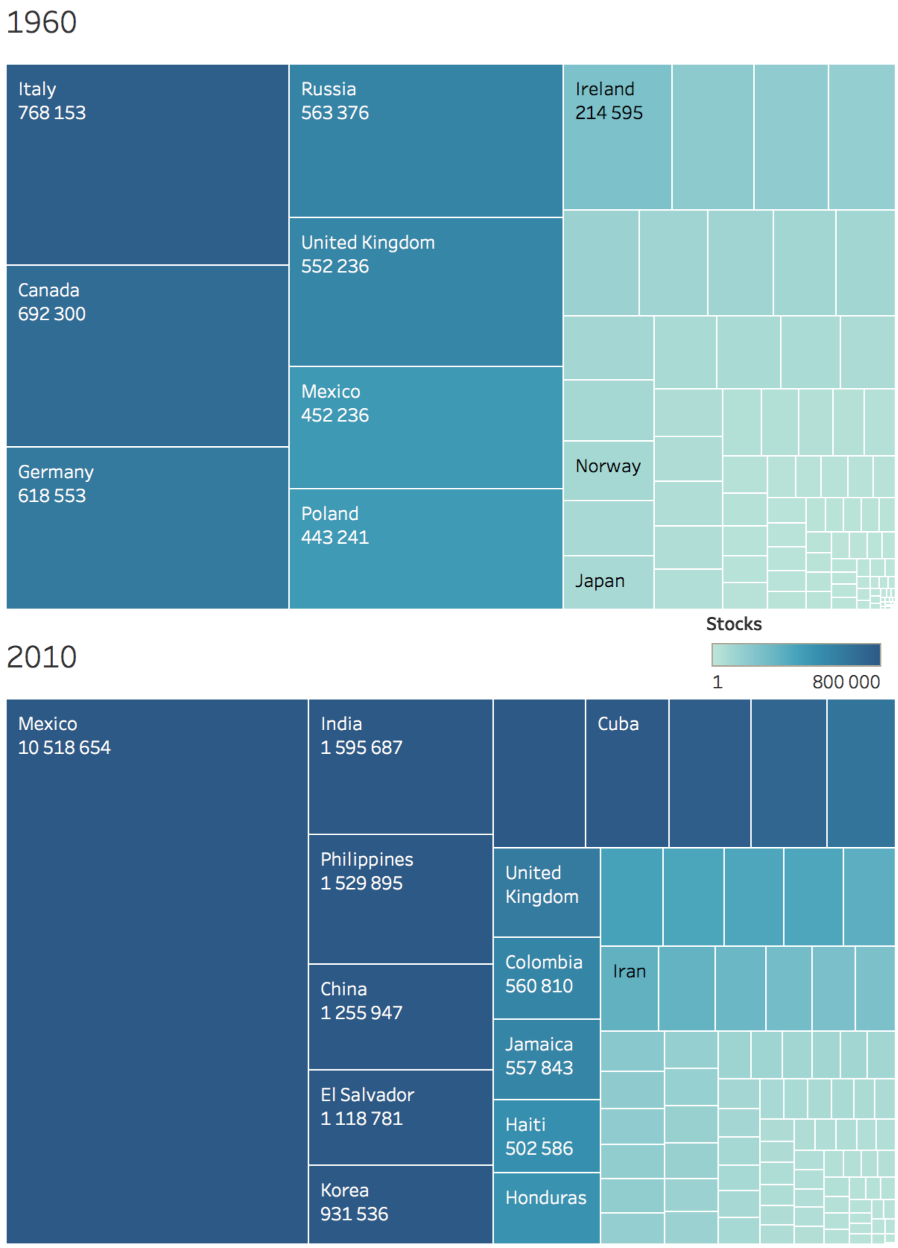
se concentre sur les 50 états et le district de Washington aux États-Unis de 1960 à 2010. Le choix de cette période est guidé par « l'Immigration and Nationality Act », qui donna lieu aux États-Unis, après son implémentation en 1965, à une augmentation considérable de l'afflux d'immigrés comme reporté dans la Figure 1.8. Ainsi, le taux d'immigration aux États-Unis est passé de 5,4 à 13,6% entre 1960 et 2010. Si l'afflux de migrants aux États-Unis sur cette période a mécaniquement entrainé une augmentation de la diversité dans la population totale, cette augmentation n'est pas visible lorsque l'on s'intéresse uniquement au groupe des migrants. Comme le montre la Figure 1.9, la probabilité de tirer deux individus nés dans deux pays différents au sein du groupe des migrants n'a pas été majoritairement affectée durant cette période, au niveau global, et ce, malgré des changements importants dans la composition des flux de migrants et notamment l'émergence de l'immigration en provenance du Mexique, des pays Asiatiques, et d'Amérique du Sud, au profit de l'immigration Européenne historique (Figures 1.8 et 1.9). Il est cependant important de noter que ces nouveaux flux de migrants ne se sont pas répartis de manière aléatoire sur tout le territoire Américain. Alors que l'afflux d'immigrés Mexicains en Californie ou au Nouveau Mexique a induit une forte baisse dans la diversité dans le groupe des migrants, des États comme le Maine ou encore le Vermont par exemple ont vu leur niveau de diversité s'accroitre de manière substantielle. Ces différences dans la répartition des nouveaux flux de migrants sont à l'origine des différences observées dans l'évolution de la diversité entre les États. Le Chapitre 2 exploite ces variations, reportées dans la Figure 1.10, afin d'étudier l'effet de la diversité sur la croissance économique[11].

L'emploi de données de panel (et donc d'effets fixes états et années) nous

ted Public Use Microdata Series).

[11] Une inquiétude légitime à propos du cas spécifique des États-Unis serait que la part importante des Mexicains dans l'immigration totale recréerait les conditions d'une corrélation forte entre l'indice de diversité et le taux d'immigration, et ce, même lorsque l'indice de diversité est calculé pour le seul groupe des migrants. Afin de réduire ces inquiétudes, nous montrons dans le Chapitre 2 que nos résultats restent inchangés en excluant la migration Mexicaine ou encore, en introduisant dans nos régressions la part que représentent les Mexicains dans l'immigration totale de chaque État. De plus, une fois les effets fixes années et États, ainsi que la part des migrants Mexicains dans l'immigration totale de chaque État, absorbés, la corrélation entre l'indice de diversité et le taux d'immigration reste faible et inférieure à 0.2.

FIGURE 1.9: Composition de l'immigration aux États-Unis entre 1960 et 2010



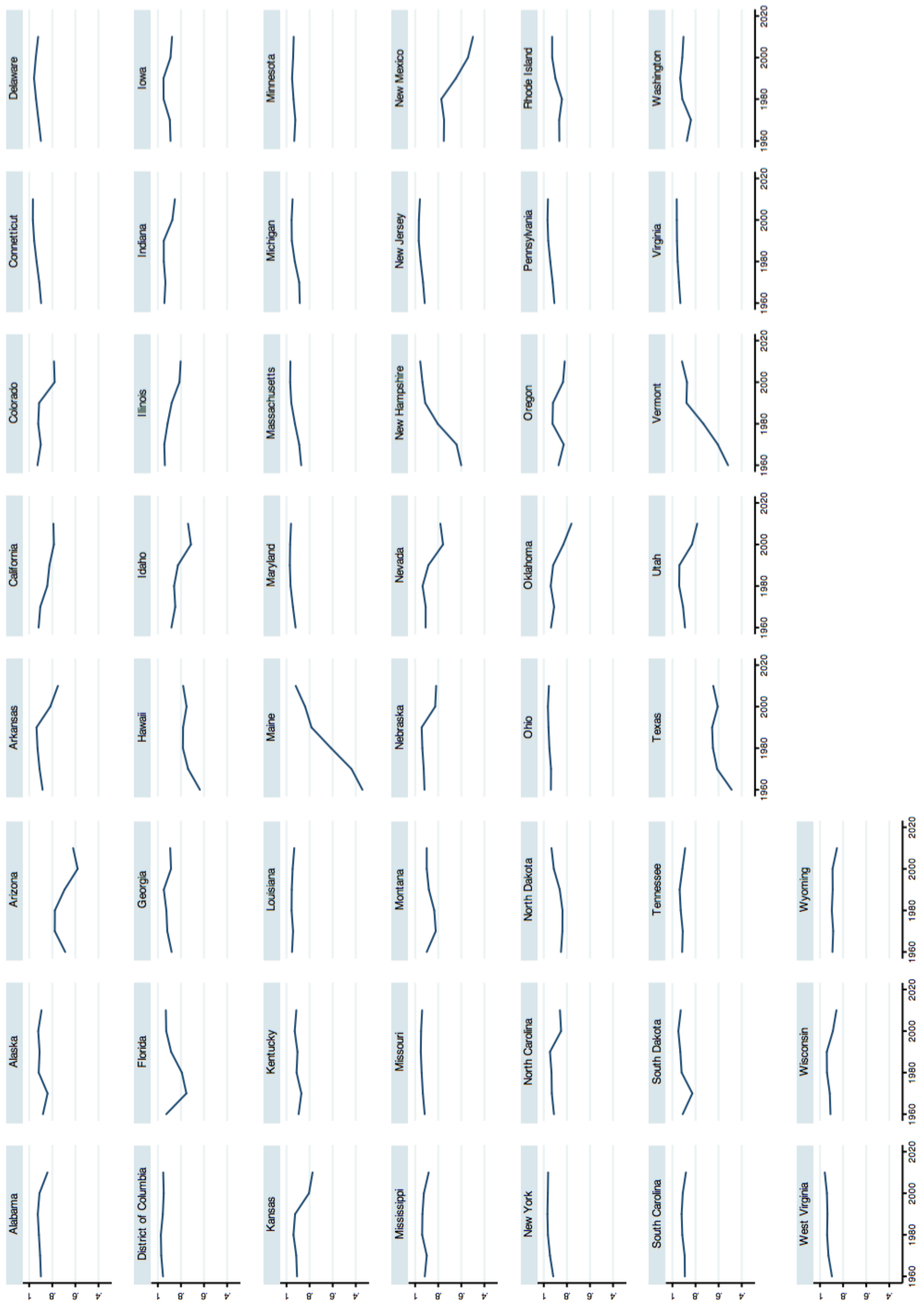
Source : Calculs de l'auteur à partir des données IPUMS.

permet de mieux isoler l'impact de la diversité en contrôlant pour l'hétérogénéité inobservée. Nous nous attachons à systématiquement étudier l'effet de la diversité en séparent les diplômés du tertiaire, des immigrés avec des niveaux d'éducation plus faibles. Aussi, nous utilisons la dimension micro-économique de nos données pour étudier l'effet de la diversité en fonction de l'âge d'entrée des migrants ou encore de leur statut légal. Cette dimension micro-économique nous permet aussi de tester dans quelle mesure le problème de causalité inverse est à même de tirer l'ensemble de nos résultats. Pour ce faire, nous identifions dans les données l'état de naissance et de résidence de chaque citoyen Américain. Nous calculons ainsi un taux de migration interne ainsi qu'un indice de diversité par état de naissance des natifs. Les effets non significatifs de la diversité, calculée comme tel, réduisent alors fortement les inquiétudes à propos d'un biais de causalité inverse. Le test placebo n'étant cependant pas une garantie parfaite, nous utilisons deux indices de diversité prédits comme instruments afin de traiter le problème d'endogénéité. Le premier indice de diversité est prédit en utilisant la répartition des migrants entre les différents États en 1960 (shift-share strategy). Le deuxième indice, quand à lui, est calculé à partir des prédictions d'un modèle de gravité à la Feyrer (2009). Nos résultats montrent que la diversité chez les éduqués du tertiaire joue de manière positive sur la croissance économique. Une augmentation de 10% de l'indice de diversité est associée à une augmentation de 6,2% du PIB par travailleur. La magnitude de cet effet est cependant à nuancer quand on sait qu'il implique que, si tous les états Américains avaient le même niveau de diversité que le District de Columbia (le niveau le plus haut de diversité observé aux USA) alors, les États-Unis n'auraient un PIB par travailleurs que 2,33% plus élevé. De plus, aucun effet significatif pour la diversité chez les migrants avec des niveaux d'éducation plus faibles ne ressort de nos estimations.

Au delà des effets purement liés à l'augmentation de la diversité dans les pays d'accueil, le débat économique s'est aussi récemment tourné vers les effets pervers que pourrait avoir l'accueil de migrants venant de pays présentant de faibles niveaux institutionnels (Collier, 2013; Borjas, 2015). L'accent a ainsi été mis dans de récents ouvrages sur l'existence de potentiels « effets de contamination » des migrants. S'il est certain que les migrants apportent avec eux les institutions prévalant dans leur pays d'origine (Fisman et Miguel, 2007), aucune étude, à ce jour, n'a clairement mis en évidence, à notre connaissance, d'effets de contamination vers les natifs. Notre étude s'attache donc à étudier ces effets de contamination en les introduisant simultanément à l'indice de diversité et au taux d'immigration. Nos résultats ne supportent en aucun cas ces effets de contamination, que cela soit au niveau des éduqués du tertiaire, ou, auprès des immigrés les moins qualifiés. Au contraire, il semble qu'attirer des immigrés de pays culturellement distants soit favorable à la croissance économique.

En conclusion, il semble, au vue de la littérature existante et de l'étude menée dans le Chapitre 2, que ni le stock de migrants ni la diversité n'aient d'impacts majeurs sur l'économie des pays receveurs de migrants ou, que les effets de l'immigration, s'ils existent, soient plutôt faiblement positifs et tirés par les diplômés du tertiaire. De manière surprenante, ces résultats semblent être en contradiction totale avec les perceptions négatives qu'une grande part des natifs des pays de l'OCDE expriment envers l'immigration. Si cette contradiction pourrait poser la question de la validité des travaux empiriques menés par les économistes, de futures recherches devront s'attacher à comprendre pourquoi les natifs semblent surévaluer les effets de la migration sur leurs économies. Il est bien sûr possible que les perceptions des natifs soient erronées comme le suggère la surestimation des taux d'immigration dans les enquêtes d'opinions « European Social Surveys » présentées dans le Tableau 1.1.

FIGURE 1.10: Diversité parmi le groupe des immigrés ente 1960 et 2010



Source : Calculs de l'auteur à partir des données IPUMS.

Table 1.1: Perceptions des natifs sur les taux d'immigration en 2015

	Taux supposé	Taux observé	Différence
Austria	26,90	17,50	9,40
Belgium	29,41	12,30	17,11
Czech Republic	9,08	2,80	6,28
Denmark	13,62	10,10	3,52
Estonia	21,88	15,40	6,48
Finland	9,89	5,70	4,19
France	26,00	12,10	13,90
Germany	22,69	14,90	7,79
Hungary	11,20	4,60	6,60
Ireland	20,35	15,90	4,45
Israel	34,34	24,90	9,44
Lithuania	11,13	4,70	6,43
Netherlands	23,78	11,70	12,08
Norway	16,35	14,20	2,15
Poland	9,19	1,60	7,59
Portugal	24,66	8,10	16,56
Slovenia	23,25	11,40	11,85
Spain	21,90	12,70	9,20
Sweden	20,80	16,80	4,00
Switzerland	31,24	29,40	1,84
United Kingdom	27,02	13,20	13,82

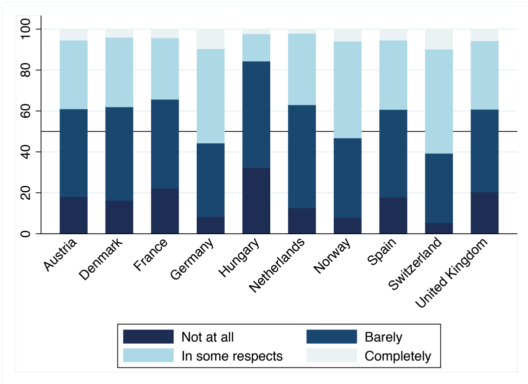
Source : Calculs de l'auteur à partie des European Social Surveys (ESS) et de la base "Trends in international migrant stocks The 2015 revision" de la division population des Nations Unies. La question posée dans les European Social Surveys est la suivante : Selon vous, sur 100 personnes vivant en [pays], combien sont nées dans un autre pays ?

1.4 De l'intégration des migrants dans leur pays d'accueil

Des attitudes largement négatives

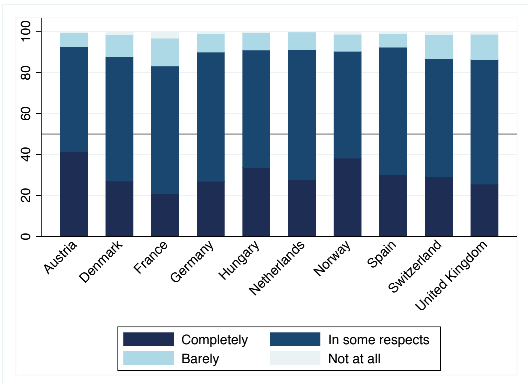
Quelles soient justifiées ou non, les perceptions des natifs et, de manière plus générale, l'attitude de ces derniers face à l'immigration, conditionnent probablement l'impact de l'immigration dans les pays de l'OCDE ; notamment au travers de l'intégration des immigrés dans la vie économique du pays. Les Figures 1.11 à 1.13 présentent une sélection des résultats des enquêtes « European Social Survey » menées dans différents pays Européens. Malgré des disparités apparentes entre pays, les attitudes des natifs envers l'immigration apparaissent clairement négatives. En moyenne, plus de 50% des répondants pensent que l'immigration n'est pas ou peu souhaitable pour l'économie ou encore que les

Figure 1.11: L'immigration est-elle bonne pour l'économie ?



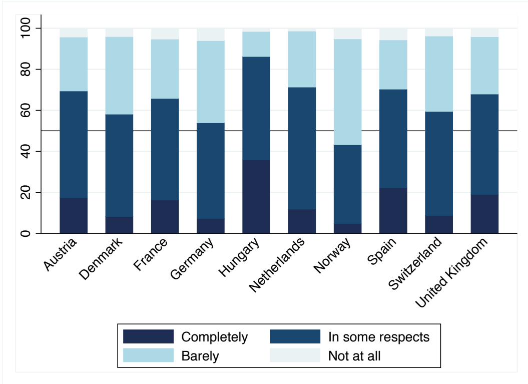
Source : Calculs de l'auteur à partir des enquêtes European Social Survey (ESS). La question posée est la suivante : Diriez-vous, en général, qu'il est plutôt bon ou mauvais que des personnes originaire d'autres pays viennent vivre ici ? Il est demandé aux individus de reporter un score allant de 0, pour « mauvais », à 10 pour « bon » pour l'économie. Nous reclassons les réponses en 4 catégories : « Not at all » (0 à 2), « Barely » (3 à 5), « In some respects » (6 à 8) et « Completely » (9 à 10). Les enquêtes European Social Survey (ESS) sont réalisées pour l'année 2014 au niveau national auprès d'un échantillon représentatif. Les observations sont ici pondérées pour prendre en compte les inégales probabilités de tirage dues au design de l'enquête.

Figure 1.12: L'immigration augmente-t-elle la criminalité ?



Source : Calculs de l'auteur à partir des enquêtes European Social Survey (ESS). La question posée est la suivante : Pensez-vous que l'immigration rende les problèmes liés aux crimes plus ou moins importants ? Il est demandé aux individus de reporter un score allant de 0, pour « augmente le crime », à 10 pour « diminue le crime ». Nous reclassons les réponses en 4 catégories : « Completely » (0 à 2), « In some respects » (3 à 5), « Barely » (6 à 8) et « Not at all » (9 à 10). Les enquêtes European Social Survey (ESS) sont réalisées pour l'année 2014 au niveau national auprès d'un échantillon représentatif. Les observations sont ici pondérées pour prendre en compte les inégales probabilités de tirage dues au design de l'enquête.

Figure 1.13: Les immigrés prennent-ils les emplois des natifs ?



Source : Calculs de l'auteur à partir des enquêtes European Social Survey (ESS). La question posée est la suivante : Pensez-vous que les immigrés prennent les emplois des natifs ou en créent de nouveaux ? Il est demandé aux individus de reporter un score allant de 0, pour « prennent les emplois », à 10 pour « créent des emplois ». Nous reclassons les réponses en 4 catégories : « Completely » (0 à 2), « In some respects » (3 à 5), « Barely » (6 à 8) et « Not at all » (9 à 10). Les enquêtes European Social Survey (ESS) sont réalisées pour l'année 2014 au niveau national auprès d'un échantillon représentatif. Les observations sont ici pondérées pour prendre en compte les inégales probabilités de tirage dues au design de l'enquête.

immigrés prennent les emplois des natifs. Ces perceptions sont encore plus mauvaises lorsque l'on s'intéresse à la criminalité puisque, en moyenne, plus de 80% des natifs déclarent que l'immigration augmente la criminalité dans leur pays. Ces attitudes donc, qu'elles soient le fruit de comportements rationnels (discrimination statistique à la Phelps 1972; Arrow 1973) ou, le fait de simples préjugés (discrimination de goût à la Becker 1957), peuvent possiblement entraver l'intégration des migrants sur le marché du travail et/ou freiner leur assimilation économique. D'une manière théorique et, dans le cadre d'un simple modèle de recherche d'emploi, la discrimination se traduit par une probabilité plus faible de recevoir des offres de travail et donc par des durées de chômage plus longues, ou encore des salaires plus faibles[12].

Une littérature à sens unique

Si la discrimination peut renvoyer au genre, à l'âge ou à la localisation géographique par exemple, la présente thèse s'intéresse principalement à la

[12] Voir Lang et Lehmann (2012) pour une revue de la littérature exhaustive sur le lien entre discrimination et performance sur le marché du travail.

discrimination ethnique, plus à même de toucher les migrants dans leur pays d'accueil. Les évidences empiriques dans cette littérature montrent bien que les écarts de salaires ne rendent pas uniquement compte de différences dans les niveaux de productivité mais bien de discriminations envers certaines minorités (Baker et Benjamin, 1994; Gundel et Peters, 2007; Charles et Guryan, 2008; Zibrowius, 2012; Kaas et Manger, 2012; Friebel et al., 2013; Biavaschi et al., 2013; Gould et Klor, 2015; De Coulon et al., 2016). Un pan particulièrement convainquant de la littérature sur les effets de la discrimination sur le marché du travail s'appuie sur des expériences aléatoires. Le papier de Bertrand et Mullainathan (2004) fait figure, dans ce cadre, de référence. En effet, Bertrand et Mullainathan (2004) étudient les différences de probabilité de rappel dans le cas de Curriculum Vitæ envoyés en réponse à des offres de travail dans les villes de Chicago et Boston. En assignant, de manière aléatoire, des noms Afro-Américains à des CV identiques en tout point de vue, Bertrand et Mullainathan (2004) montrent que les employeurs tendent à privilégier les « blancs » par rapport à ces derniers. Les « blancs » ont ainsi 50% de chances supplémentaires de recevoir une réponse par rapport aux individus avec un nom Afro-Américain.

Contribution de la thèse : De l'hétérogénéité de la discrimination

S'il apparait alors évident que des migrants peuvent se trouver discriminés en raison de leur appartenance à une ethnie ou à une origine particulière, et que cela peut entraver leurs performances sur le marché du travail, peu de papiers s'intéressent à l'hétérogénéité observée dans les niveaux de discrimination. Or, tous les immigrés ne font pas nécessairement face aux mêmes attitudes de la part des natifs et ces différences peuvent être une des explications possibles des écarts de performance observés pour des migrants d'origines différentes dans un même pays d'accueil. Le Chapitre 3 de cette thèse se donne alors pour objectif d'examiner empiriquement cette question. Pour ce faire, nous avons recours à des données de panel au niveau individuel sur la période 1984-2012 (German Socio Economic Panel). Ces données nous permettent de reconstruire la trajectoire de chaque migrant sur le marché du travail. Pour chaque mois nous savons si l'individu était employé, au chômage, ou inactif. A l'aide d'un

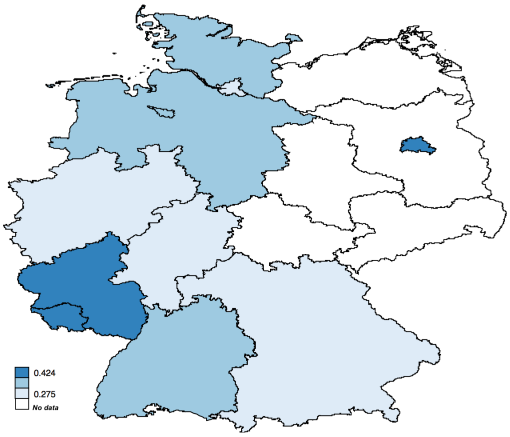
TABLE 1.2: Confiance des Allemands envers les citoyens des principaux pays d'origine des migrants

Country of origin	Trust (%)
Austria	80,0
Netherlands	71,1
France	68,9
United States	68,5
Spain	66,2
Portugal	59,7
Greece	59,6
United Kingdom	56,5
Italy	55,8
Hungary	51,7
Czech Republic	39,2
Russia	34,7
Turkey	32,0
Poland	26,8
Romania	24,4

Source : Calculs de l'auteur à partir des enquêtes Eurobarometer et European Election Survey data. Seul les origines utilisées dans le Chapitre 3 sont reportées. Trust représente le % d'Allemands qui déclarent avoir confiance dans les citoyens du pays en question.

modèle de durée, nous estimons alors l'influence de l'attitude des natifs envers chaque pays d'origine des migrants sur leurs durées de chômage respectives. Les données sur les attitudes sont issues des enquêtes « Eurobarometers » et « European Election Surveys » et mesurent la confiance des Allemands envers les citoyens des différents pays d'origine des migrants. La confiance des Allemands, au niveau national, comme en témoigne le Tableau 1.2 est très dépendante du pays d'origine des migrants et liée à des différences culturelles, politiques et institutionnelles (Guiso *et al.*, 2009). Cette confiance varie aussi, pour une origine donnée, entre les différentes régions Allemandes, et ce, en raison de différences culturelles notamment. La Figure 1.14 montre bien cela en reportant ces variations régionales pour les Turcs, la diaspora la plus importante en Allemagne avec plus de 1 655 000 migrants. Le Chapitre 3 de cette thèse utilise ces variations au niveau national et régional pour estimer l'impact de la confiance des natifs sur l'intégration des migrants sur le marché du travail.

FIGURE 1.14: Confiance des Allemands envers les Turcs.



Source : Calculs de l'auteur à partir des enquêtes Eurobaromaters European Election Survey.

Nos résultats montrent que des niveaux de confiance plus faibles des natifs envers les résidents d'un pays donné sont associés à des durées de chômage plus longues pour les immigrés originaires de ce dernier. Concernant la magnitude de cet effet, si les Allemands déclaraient les mêmes niveaux de confiance envers les Turcs qu'envers les Autrichiens, alors les immigrés Turcs verraient leurs durées de chômage réduites de 3 mois en moyenne. Au delà des implications politiques d'un tel résultat, le Chapitre 3 souligne l'importance de dépasser les problèmes d'identification dus à l'auto-sélection des migrants. En effet, le problème majeur de l'analyse d'un stock auto-sélectionné de migrants est que ce dernier exclu par définition les individus pour lesquels la discrimination est la plus couteuse ; entrainant mécaniquement un biais à la baisse dans l'effet estimé de la discrimination sur l'emploi. L'ajout d'effets fixes origines-années dans notre modèle, et, l'utilisation de données au niveau régional, nous permettent de dépasser ce problème comme détaillé dans le Chapitre 3.

1.5 De l'impact des migrants sur leur pays d'origine

Jusqu'à présent cette introduction s'est uniquement concentrée sur deux des trois groupes concernés par la migration internationale, à savoir, les migrants eux-mêmes et les natifs dans leur pays d'accueil. Or, si ces deux groupes occupent une place centrale dans le débat sur les enjeux de l'immigration dans les pays développés, la migration internationale affecte en premier lieu les pays de départ de migrants et leurs proches restés dans leur pays d'origine.

La littérature sur l'impact de la migration sur les pays en développement s'est polarisée autour de trois axes à savoir (i) les externalités de la diaspora sur l'investissement en capital humain dans les pays d'origine des migrants (Mountford, 1997; Stark *et al.*, 1997; Beine *et al.*, 2001, 2008; Docquier et Rapoport, 2012; Gibson et McKenzie, 2012) (ii) les effets des envois de fonds des migrants[13] (Docquier et Rapoport, 2006; Yang, 2011; Combes et Ebeke, 2011; Brown et Jimenez-Soto, 2015) (iii) les transferts intangibles des diasporas vers leur pays d'origine. Si la fuite des cerveaux ou les envois de fonds influencent de manière substantielle les pays en développement, la présente thèse s'intéresse plutôt au troisième axe : les transferts intangibles des migrants vers leur pays d'origine. Ces transferts non monétaires peuvent être de plusieurs sortes et affecter des dimensions aussi diverses que la fertilité (Beine *et al.*, 2013; Bertoli et Marchetta, 2015), la discrimination envers les femmes (Lodigiani et Salomone, 2015; Tuccio et Wahba, 2016) ou encore la démocratie (Spilimbergo, 2009; Mercier, 2016; Docquier *et al.*, 2016; Barsbai *et al.*, 2017), pour ne citer qu'eux .

Contribution de la thèse : Au delà des transferts produit à produit

Le Chapitre 4 de cette thèse s'intéresse spécifiquement aux transferts de technologies et de connaissances que peuvent engendrer les migrations internationales. En effet, en migrant vers des pays à haut revenu, les migrants se trouvent en contact direct avec des technologies avancées qu'ils peuvent alors transmettre vers leur pays d'origine. Ces transferts peuvent, entre autre, se faire par la migration retour ou encore la baisse des couts commerciaux et de

[13] Selon la Banque mondiale en 2015 les envois de fonds de migrants vers les pays en développement ont atteint la somme de 440 milliards de dollars. Cela en fait le deuxième flux financier au niveau international derrière les IDE mais devant l'aide au développement ou encore les portefeuilles privés de participations. La présente thèse ne s'intéresse pas en soit à l'effet des envois de fonds au niveau international qui constitue une autre littérature à lui seule.

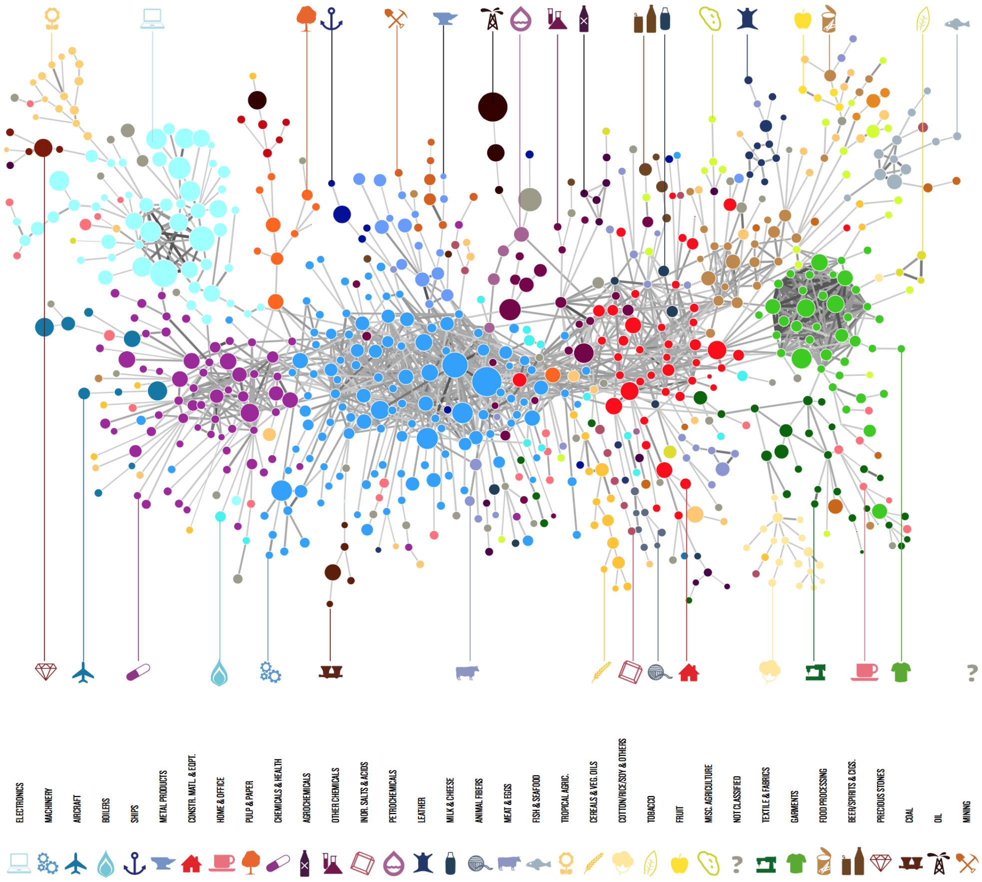
communication par exemple. S'il est vrai que ce sujet a déjà fait l'objet d'une attention toute particulière dans la littérature[14] (Kerr, 2008; Lodigiani, 2008; Mayr et Peri, 2009; Andersen et Dalgaard, 2011; Agrawal *et al.*, 2011; Kerr et Kerr, 2015; Naghavi et Strozzi, 2015; Breschi *et al.*, 2017), la plupart des papiers se sont concentrés sur des indicateurs de technologies et de productivité avancés, comme les brevets par exemple. Or, si ces indicateurs rendent compte de l'évolution de la partie haute des arbres technologiques, ils ne représentent qu'une partie de l'ensemble des connaissances transférées par les migrants vers leur pays d'origine. Les travaux de Bahar et Rapoport (2016) font exception à cette règle en utilisant les données commerciales au niveau international[15]. Avec une analyse très fine au niveau produit, Bahar et Rapoport (2016) montrent comment la migration est un déterminant majeur de l'évolution des avantages comparatifs à travers les transferts technologiques de la diaspora. Bahar et Rapoport (2016) estiment qu'une augmentation de l'émigration pour un pays donné équivalente à 65 000 individus est associée à une probabilité plus forte de 15% de développer un avantage comparatif dans un des biens exportés par le pays d'accueil des migrants. Ils démontrent ainsi que les migrations internationales façonnent le panier d'exportations des pays et ce d'une façon plus importante que le commerce ou encore les IDE. Ces auteurs ne font donc aucune restriction sur les types de technologies pouvant être transférées par les migrants. En revanche, l'analyse au niveau produit n'autorise les transferts technologiques que d'un produit vers celui-ci. Cependant, nous savons que la technologie employée pour un produit peut servir à la production de beaucoup d'autres. Les connaissances en mécanique nécessaires à la production de voitures peuvent être employées dans la production d'autres biens comme les motos par exemple. « L'espace produits » (Figure 1.15) résume cette idée selon laquelle les connaissances requises pour la production d'un bien sont employées

[14] L'ensemble des papiers attenant à cette littérature sont détaillés dans l'introduction du Chapitre 4.

[15] Dans un papier plus récent, Bahar et al. (2017) utilisent les migrations retours de centaines de milliers de migrants Yougoslaves en Allemagne entre 1995 et 2000 comme une expérience naturelle pour évaluer l'impact de la migration sur l'évolution des avantages comparatifs des pays d'origine des migrants. Ces auteurs trouvent que l'élasticité des exportations aux migrations retours varie entre 0.1 et 0.25 dans les industries où les migrants étaient employés durant leur séjour en Allemagne.

dans la production d'autres bien. Chaque point dans cet espace représente un produit exporté sur le marché international quand chaque lien connecte les produits ayant une forte probabilité d'être co-exportés. Cette visualisation des exportations mondiales montre bien que les produits ont tendance à s'agglomérer au sein de groupes fortement connectés entre eux. Ces groupes rassemblent des produits qui requièrent des connaissances productives très proches voire identiques. Les diasporas installées dans des pays à fort potentiel technologique représentent donc une formidable opportunité pour les pays d'origine des migrants d'acquérir les connaissances nécessaire à la production d'un bien nouveau. Ensuite, et, à partir de ces nouvelles connaissances, ces pays peuvent se déplacer le long de « l'espace produits » en diversifiant leur production. Le Chapitre 4 de cette thèse prend alors en compte cette proximité entre les produits pour estimer un effet au delà de celui étudié par Bahar et Rapoport (2016). Pour ce faire nous utilisons l'ECI (Economic Complexity Index, Hausmann *et al.* (2011)), un indicateur agrégé de la technologie et plus largement des connaissances productives présentes dans un pays. Nous empruntons notre spécification à la littérature sur les transferts de normes et mettons en évidence que l'augmentation du stock de migrants dans 2O pays de l'OCDE est associée à une augmentation de l'ECI dans les pays d'envoi des migrants. Nous utilisons un modèle de panel dynamique afin de tenir compte, non seulement des effets d'inertie dans la technologie, mais aussi, afin de différencier les effets de court terme et de long terme de la migration. L'endogénéité est traitée en utilisant l'estimateur des « System GMM », à la fois avec instruments internes et externes. Nos instruments externes sont construits à partir des prédictions d'un modèle de gravité à la Feyrer (2009). Dans notre modèle principal, nos estimations impliquent que si aucun migrant n'avait quitté le Mexique en 2005 alors l'ECI de ce dernier aurait été plus faible de 0.197 par rapport à la valeur réelle observée pour l'année 2010. Le Mexique aurait alors perdu 3 places dans le classement international de l'ECI en 2010 ; et cet effet aurait perduré à long terme représentant alors une perte de 25 places au total. Les résultats obtenus dans le Chapitre 4 sont robustes à la prise en compte des autres canaux de transmissions de la technologie au niveau international comme le commerce ou les IDE et à l'emploi de différents indicateurs de technologie. Aussi, nous

FIGURE 1.15: L'espace produits



Source : Hausmann *et al.* (2011).

mettons en évidence le fait que les transferts technologiques sont d'autant plus grand que le pays d'accueil des migrants est technologiquement avancé et que l'émigration est importante.

Bibliographie

Ager, P., and Brückner, M. (2013). Cultural diversity and economic growth: Evidence form the US during the age of mass migration. *European Economic Review*, 64:76–97.

Alesina, A., Harnoss, J. and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2):101–138.

Andersen, T. B. and Dalgaard, C. J. (2011). Flows of people, flows of ideas, and the inequality of nations. *Journal of Economic Growth*, 16(1):1–32.

Angelucci, M. (2015). Migration and financial fonstraints: Evidence from Mexico. *The Review of Economics and Statistics*, 97(1):224–228.

Agrawal, A., Kapur, D. McHale, J. and 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.

Arrow, K. (1973). The theory of discrimination in: O. A. Ashenfelter and A. Rees, eds., *Discrimination in labor markets.* Princeton, N.J: Princeton University Press, 3–33.

Bahar, D. and Rapoport, H. (2016). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal.*, forthcoming.

Bahar, D., Hauptmann, A., Özgüzel C., and Rapoport, H. (2017). Diasporas, return migration and comparative advantage: A natural experiment of Yugoslavian refugees in Germany. Presented at the 10th AFD-World Bank Migration and Development Conference, Clermont-Ferrand, France.

Baker, M. and Benjamin, D. (1994). The performance of immigrants in the Canadian labor market. *Journal of Labor Economics*, 12(1):369–405.

Barbone, L., Bontch-Osmolovsky, M., and Zaidi, S. (2009). The foreign-born population in the European Union and its contribution to national tax and benefit systems. World Bank Policy Reasearch Working Paper No. 4899.

Barsbai T., Rapoport, H., Steinmayr, A., and Trebesch, C. (2017). The effect of labor migration on the diffusion of democracy: Evidence from a former Soviet Republic. *American Economic Journal: Applied Economics.*, forthcoming.

Becker, G. S. (1957). *The economics of discrimination.* Chicago: the University of Chicago Press.

Beine, M., Docquier, F., and Schiff, M. (2013). International migration, transfer of norms and home country fertility. *Canadian Journal of Economics*, 46(4):1406–1430.

Beine, M., Docquier, F., and Rapoport, H. (2001). Brain drain and economic growth: Theory and evidence. *Journal of Development Economics*, 64(1):275–289.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain drain and human capital formation in developing countries: Winners and losers. *Economic Journal*, 118(528):631–652.

Bertoli, S. and Fernandez-Huertas Moraga J. (2015). The size of the cliff at the border. *Regional Science and Urban Economics*, 51(1):1–6.

Bertoli, S. and Marchetta, F. (2015). Bringing it all back home–return migration and fertility choices. *World Development*, 65:27–40.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94(4):991–1013.

Biavaschi, C., Giulietti, C., and Siddique, Z. (2013). The economic payoff of name Americanization. *Journal of Labor Economics.*, forthcoming.

Boeheim, R., Horvath, G., and Mayr, K. (2012). Birthplace diversity of the workforce and productivity spill-overs in firms. WIFO Working PapersNo. 438.

Borjas, G.J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics* , 118(4):1335–1374.

Borjas, G.J. (2015). Immigration and globalization: A review essay. *Journal of Economic Literature*, 53(4):961–974.

Borjas, G.J. (2017). The wage impact of the *Marielitos*: A reappraisal. *Industrial and Labor Relations Review*, *forthcoming.*

Borjas, G., J., Freeman, R., B., and Katz, L., F. (1997). How much do immigration and trade affect labor market outcomes?. Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 28(1), pages 1-90.

Borjas, G., J., Grogger, H., H., and Hanson, G., H. (2012). Comment: On estimating elasticities of substition. *Journal of the European Economic Association*, 10(1):198–210.

Breschi, S., Lissoni, F., and Miguelez, E. (2017). Foreign-origin inventors in the USA: Testing for diaspora and brain gain effects. *Journal of Economic Geography.* doi: https://doi.org/10.1093/jeg/lbw044

Brown, R., P. and Jimenez-Soto, E. (2015). Migration and remittances. in *Handbook of the economic of international migration*, 1: 1077–1140. Edited by Barry R. Chiswick and Paul W. Miller, North-Holland.

Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *Industrial and Labor Relations Review*, 43(2):245–257.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Carrington, W., J. and Lima, J., F. (1996). The impact of 1970s repatriates from Africa on the Portuguese labor market. *Industrial and Labor Relations Review*, 49(2):330–347.

Charles, K., K., and Guryan, J. (2008). Prejudices and wages: An empirical assessment of Becker's The Economics of Discrimination.. *Journal of Political Economy*, 116(5):773–809.

Clemens, M., A., Montenegro C., E., and Pritchett, L. (2010). The place premium: Wage differences for identical workers across the US border. Series Documentos de Trabajo No. 314, University of Chile.

Clemens, M., A. (2011). Economics and emigration: Trillion-dollar bills on the sidewalk? *Journal of Economic Perspectives*, 25(3):83–106.

Clemens M. (2013). The effect of foreign labor on native employment. A job-specific approach and application to North Carolina farms. Center for Global Development Working Paper No. 326.

Collier, P. (2013). *Exodus: How migration is changing our World.* Oxford University Press.

Combes, J-L. and Ebeke, C. (2011). Remittances and household consumption instability in developing countries. *World Development*, 39(7):1076–1089.

Cortes, P. (2008). The effect of low-skilled immigration on U.S. prices: Evidence from CPI data. *Journal of Political Economy*, 116(3):381–442.

De Coulon, A., Radu, D., and Friedrich Steinhardt, M. (2016). Pane e cioccolata: The impact of native attitudes on return migration. *Review of International Economics*, 24(2):253–281.

Docquier, F. and Marfouk, A. (2006). International migration by educational attainment, 1990-2000. in *International migration, Remittances and Development.* In: Özden, Ç., and Schiff M. vol. 5. Palgrave Macmillan: New York (2006). p. 151–199.

Docquier, F. and Rapoport, H. (2006). The economics of migrants' remittances. in *Handbook of the economics of giving, altrusim and reciprocity*, Vol.2, p. 1135–1198. Edited by Serge-Christophe Kolm and Jean Mercier Ythier.

Docquier, F. and Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of Economic Literature*, 50(3):681–730.

Docquier, F., Machado, J., and Sekkat, K. (2015). Efficiency Gains from liberalizing labor mobility. *The Scandinavian Journal of Economics*, 117(2):303–346.

Docquier, F., Lodigiani, E., Rapoport, H., and Schiff, M. (2016). Emigration and democracy. *Journal of Development Economics*, 120(1):209–223.

Feyrer, J. (2009). Trade and income: Exploiting time series in geography. NBER working paper No. 14910. National Bureau of Economic Research Cambridge, MA.

Fisman, R., and miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political Economy*, 115(6):1020–48.

Friebel, G., Gallego, J. M., and Mendola, M. (2013). Xenophobic attacks, migration intentions, and networks: Evidence from the South of Africa. *Journal of Population Economics*, 26(2):555–591.

Friedberg, R., M. (2001). The impact of mass migration on the Israeli labor market. *The Quarterly Journal of Economics*, 116(4):1373–1408.

Friedberg, R., m., and Hunt, J. (1995). The impact of immigrants on host country wages, employment and growth. *Journal of Economic Perspectives*, 9(2):23–24.

Fulford, S., Petkov, I., and Schiantarelli, F. (2015). Does it matter where you came from? Ancestry composition and economic performance of U.S. counties, 1850-2010. Boston College Working Papers in Economics No. 875.

Gibson, J. and McKenzie, D. (2012). The economic consequences of 'Brain Drain' of the best and brightest: Microeconomic evidence from five countries. *Economic Journal*, 122(560):339–375.

Gould, E. D. and Klor, E. F. (2015). The long-run effect of 9/11: Terrorism, backlash, and the assimilation of Muslim immigrants in the West. *The Economic Journal (forthcoming).*

Guiso, L., Sapienza, P., and Zingales, L. (2009). Cultural biases in economic exchange?. *Quarterly Journal of Economics*, 124(3):1095–1131.

Gundel, S. and Peters, H. (2007). Assimilation and cohort effects for German immigrants. *SOEP paper No.64.*

Hatton, T., J. and Williamson J., G. (2006). What drove European mass emigration? In Global Migration and the World Economy, by T. J. Hatton and J. G. Williamson, 51–76. Cambridge, MA: MIT Press.

Haussman, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., and Yildirim, M. A. (2011). The atlas of Economic Complexity. Puritan Press. Cambridge MA.

Hunt, J. (1992). The impact of the 1962 repatriates from Algeria on the French labor market. *Industrial and Labor Relations Review* , 45(3):556–572.

Kaas, L. and Manger, C. (2012). Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review*, 13(1):1–20.

Kerr, S. P. and Kerr, W. R. (2015). Global collaborative patents. NBER Working Paper 21735. National Bureau of Economic Research Cambridge, MA.

Kerr, W. R. (2008). Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics*, 90(3):518–537.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: Assessing the impacts of cultural and language diversity in the national hockey league. *The Review of Economics and Statistics*, 95(1):302–314.

Lang, K. and Lehmann, J-Y. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature*, 50(4): 959–1006.

Lodigiani, E. (2008). Diaspora externalities and technology diffusion. *Economie Internationale*, 115(3):43–64.

Lodigiani, E. and Salomone, S. (2015). Migration-induced transfers of norms: The case of female political empowerment. Working Papers No. 19. Department of Economics, University of Venice Ca'Foscari.

Mayr, K. and Peri, G. (2009). Brain drain and brain return: Theory and application to Eastern-Western Europe. *The BE Journal of Economic Analysis & Policy*, 9(1):1–52.

McKenzie, D. (2006). Beyond Remittances: The Effects of Migration on Mexican Households. in *International Migration, Remittances and the Brain Drain*, C., Özden and M., Shiff, The World Bank.

McKenzie, D. et Rapoport, H. (2011). Can Migration Reduce Educational Attainment ? Evidence From Mexico. *Journal of Population Economics*, 24(4):1331–1358.

Clemens, M., A., Montenegro C., E., and Pritchett, L. (2010). How important is selection? Experimental vs. Non-Experimental measures of the income gains from migration. *Journal of the European Economic Association*, 8(4):913–945.

Messias, J., Benevenuto, F., Weber, I. and Zagheni, E. (2016). From migration corridors to clusters: The value of Google+ data for migration studies. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

McKeown, A. (2004). Global migration, 1846-1940. *Journal of World History*, 15(2):155–189.

Mercier, M. (2016). The return of the prodigy son: Do return migrants make better leaders?. *Journal of Development Economics*, 122(C):76–91.

Mountford, A. (1997). Can a Brain Drain Be good for growth in the source economy? *Journal of Development Economics*, 53(2):287–303.

Naghavi, A. and Strozzi, C. (2015). Intellectual property rights, diasporas, and domestic innovation. *Journal of International Economics*, 96(1):150–161.

OECD (2013). The fiscal impact of immigration in OECD countries, in International Migration Outlook 2013, OECD Publishing, Paris.

Ottaviano, G., and Peri, G. (2006). The economic value of cultural diversity: Evidence from U.S. cities. *Journal of Economic Geography*, 6(1):9–44.

Ottaviano, G., and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Özden, Ç., Parsons C. R., Schiff, M., Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960-2000. *World Bank Economic Review*, 25(1):12–56.

Ozgen, C., Peters, C., Niebuhr, A., Nijkamp, P. and Poot, J. (2014). Does cultural diversity of migrants employees affect innovation? *International Migration Review*, 48(9):377–416.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2014). Does labor diversity affect firm productivity? *European Economic Review*, 66:144–179.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(1):659–661.

Pritchett L. (2009). The cliff at the border. R. Kanbur, M. Spence (Eds.), Equity and Growth in a Globalizing World, The World Bank and the Commission on Growth and Development, Washington (2009), pp. 263–286

Spilimbergo, A. (2009). Democracy and foreign education. *The American Economic Review*, 99(1):528–543.

Suedekum, J., Wolf, K., and Blien, U. (2014). Cultural diversity and local labour markets. *Regional Studies*, 48(1):173–191.

Stark, O., Helmenstein, C., and Prskawetz, A. (1997). A Brain Gain with a Brain Drain.. *Economic Letters*, 55(2):227–234.

Tuccio, M. and Wahba, J. (2016). Can I have permission to leave the house? Return migration and the transfer of gender norms. IZA Discussion Paper 9216. Institute for the Study of Labor (IZA), Bonn, Germany.

United Nations (2015). Trends in international migrant stock: The 2015 revision. Department of Economic and Social Affairs, Population Division.

Yang, D. (2011). Migrant remittances. *The Journal of Economic Perspectives*, 25(3):129–151.

Zibrowius, M. (2012). Convergence or divergence? Immigrant wage assimilation patterns in Germany. *SOEP paper No.479.*

Chapter 2

Multiculturalism and Growth: Skill-Specific Evidence from the Post-World War II Period

This chapter is joint work with Frédéric Docquier (FNRS & IRES, Université Catholique de Louvain, Belgium), Riccardo Turati (IRES, Université Catholique de Louvain, Belgium) and Chrysovalantis Vasilakis (Bangor Business School, United-Kingdom).[1] It is currently "under review" at *Journal of Economic Growth.*

2.1 Introduction

Patterns of international migration to industrialized countries have drastically changed since World War II (WW2). On average, the share of foreigners in the population of high-income countries increased from 4.9 to 11.7% between 1960 and 2010 (Özden *et al.*, 2011).[2] This phenomenon has similarly affected the

[1]We are grateful to Simone Bertoli, Jean-Louis Combes, Oded Galor and Hillel Rapoport for helpful comments. This chapter has also benefited from discussions at the SEPIO Workshop on Cultural Diversity (Paris, May 2016), at the Workshop on "The Importance of Elites and their Demography for Knowledge and Development" (Louvain-la-Neuve, June 2016), at the XII Migration Summer School at EUI (Florence, July 2016), at the 7th International Conference "Economics of Global Interactions: New Perspectives on Trade, Factor Mobility and Development" (Bari, September 2016), at the GDRI International Development Conference (Clermont-Ferrand, November 2016), at the DIAL Seminar (Paris, January 2017) and at the 10th AFD-World Bank Migration and Development Conference (Clermont-Ferrand, June 2017).

[2]This is not the case in developing countries, where the average immigration rate has decreased by half (from 2.3 to 1.1%) since 1960. Although the worldwide stock of international migrants increased from 91.6 to 211.2 million, the worldwide share of international migrants has been fairly stable since 1960, fluctuating around 3%. This is only 0.3 percentage points above the level observed in the early 20th century (McKeown, 2004).

United States (from 5.4 to 13.6%), the members of the European Union (from 3.9 to 12.2%), Canada and Australia (from 15 to 22%). In addition, this change has been predominantly driven by immigration from developing countries; the share of South-North immigrants in the population of high-income countries increased from 2.0 to 8.7% in half a century.[3] This growing inflow of people coming from geographically, economically and culturally distant countries raises specific issues, as it has conceivably brought different skills and abilities, but also different social values and norms, or different ways of thinking. Although a large body of literature has focused on the size and skill structure of immigration flows, the macroeconomic effects of multiculturalism, as well as the channels through which they materialize, are still uncertain.

This paper empirically revisits the impact of multiculturalism on the macroeconomic performance of US states (proxied by their level of GDP per capita) in the aftermath of WW2. Our analysis combines three distinctive features. First, we rely on panel data available for a large number of regions over a long period. Our sample covers all US states over the 1960-2010 period in ten-year intervals. The use of panel data allows us to better deal with unobserved heterogeneity and endogeneity issues. This is crucial because economic prosperity and the degree of diversification of production are likely to attract people from different cultural origins. Multiculturalism is thus likely to respond to changes in the economic environment (see (Alesina and La Ferrara, 2005)), implying that causation is hard to establish in a cross-sectional setting. To control for unobserved heterogeneity and reverse causation biases, our paper uses a great variety of geographic and time fixed effects, and combines various instrumentation strategies that have been used in the existing literature. Second, we systematically investigate whether the economic effect of multiculturalism is heterogeneous across skill groups. The costs and benefits from multiculturalism are likely to vary with the levels of task complexity and interaction between workers; meanwhile, high-skilled and low-skilled immigrants are likely to heterogeneously propagate social values and norms across borders. We account

[3]Immigration from developing countries accounts for 98% of the 1960-2010 rise in immigration to high-income countries, for 80% in the European Union, for 120% in the United States, and for 150% in Australia and Canada. Trends in immigration to the US are presented in the Appendix.

for this by using skill-specific measures of multiculturalism. In addition, taking advantage of the availability of microdata, we compute our indices of multiculturalism for different groups of immigrants (by age of entry or by legal status). Third, we jointly test for different technologies of transmission. We follow Alesina *et al.* (2016) and proxy multiculturalism with indices of birthplace diversity, measuring the probability that two randomly-drawn individuals from a particular state have different countries of birth. In alternative specifications, we allow for non-linear effects, and include epidemiological (or contamination) forces, as well as an index of birthplace polarization of the workforce.

Our paper belongs to a recent and increasing strand of literature which considers that culture can be a feature which differentiates individuals in terms of their attributes, that this differentiation may have positive or negative effects on people's productivity, and that culture is affected by the country of birth (which determines the language and social norms individuals were exposed to in their youth, the education system, etc.). On the one hand, homogenous people are more likely to get along well, which implies that multiculturalism may reduce trust or increase communication, cooperation and coordination costs. Moreover, birthplace diversity can also be the source of epidemiological effects, as argued by Collier (2013) and Borjas (2015): by importing their "bad" cultural, social and institutional models, migrants from developing countries may contaminate the entire set of institutions in their country of adoption, influencing the world distribution of technological capacity. On the other hand, cultural diversity also enhances complementarities across diverse productive traits, stimulating innovations and the collective capacity to solve problems; a more diverse group is likely to spawn different cultures with various solutions to the same problem. Evidence of such costs and benefits has been found in micro studies. For example, Parrotta *et al.* (2014) investigate the effect of different forms of diversity (by education, age group, and nationality) on the productivity of Danish firms, using a matched employer-employee database. They find a negative effect of workers' diversity by nationality on productivity. On the contrary, Ozgen *et al.* (2014) find that birthplace diversity increases the likelihood of innovations using Dutch firm-level survey data, and Boeheim *et al.* (2012) find a positive effect of diversity on productivity using Austrian data. Finally, Kahane *et al.* (2013) find a positive effect of diversity on hockey team performance using data from the NHL (the North American National Hockey League).

Contrary to the firm-level approach, the analyses conducted at the macro level account for interdependencies between firms, industries, and/or regions. Existing studies have identified significant and positive effects of multiculturalism on comparative development and on disparities in economic performance across modern societies.[4] Ottaviano and Peri (2006) use US data by metropolitan area over the 1970-1990 period. In their (log of) wage regressions, the coefficient of diversity varies between 0.7 and 1.5. Ager and Brückner (2013) use US data by county during the 1870-1920 period: the coefficient of diversity in the output per capita regressions varies between 0.9 and 2.0. In these two studies, endogeneity issues are solved by using a shift-share method, i.e. computing the diversity index on the basis of predicted immigrant stocks. More precisely, the change in immigration to a region is predicted as the product of the global change in immigration to the US by the regional share in total immigration in the initial year. A more recent study accounting for the education level of immigrants is that of Alesina *et al.* (2016); it is the most similar to ours. They use cross-sectional data on immigration stocks by education level for a large set of countries in the year 2000, and develop a pseudo-gravity first-stage model to predict migration stocks and birthplace diversity indices. They also identify a positive effect of birthplace diversity in countries with GDP per capita above the median, and a stronger effect for diversity among college-educated workers. The effect of diversity on the log of GDP per capita is around 0.1 when computed on low-skilled workers, while the effect of diversity among the highly skilled varies between 0.2 and 0.3. Similarly, Suedekum *et al.* (2014) use annual German data by region from 1995 to 2006. Over this short period, they find a lower effect of diversity on the log of German wages (about 0.1 for

[4] Ashraf and Galor (2013) use the concept of genetic diversity (capturing within-group heterogeneity in genomes between regions), and find that it explains about 25% of the different development outcomes (as proxied by population density) around the year 1500, i.e. before the age of mass migration. They identify an inverted-U shape relationship, suggesting that there is an optimal level of diversity for economic development. On the contrary, the empirical literature on ethnic and linguistic fractionalization identifies negative effects on economic growth (at least in developing countries).

diversity among high-skilled foreigners, and 0.04 for diversity among the low skilled) when fixed effects and IV methods are used.

Our empirical analysis relies on high-quality US census data by state over the 1960-2010 period. The choice of this period is guided by the 1965 amendments to the Immigration and Nationality Act, which led to an upward surge in U.S. immigration and diversity (as in (Ottaviano and Peri, 2006)). Birthplace diversity is almost perfectly correlated with the state-wide proportion of immigrants, which has increased threefold since 1960 in all skill groups. It is thus statistically impossible to disentangle the effects of birthplace diversity from those of the size of immigration. For this reason, we opt for a benchmark model that includes the immigration rate and a birthplace diversity index pertaining to the immigrant population. In line with Alesina *et al.* (2016) and Suedekum *et al.* (2014), we find that diversity among college-educated immigrants is positively associated with the level of GDP per capita; however, diversity among less educated immigrants has insignificant (or weakly significant) effects. Another remarkable result is that the estimated coefficient is divided by four when geographic and year fixed effects are included. Overall, a 10% increase in high-skilled diversity raises GDP per capita by 6.2%. These results are robust to the exclusion of some census years, to the set of US states included in the sample, to the measurement of diversity, and to the definition of a high-skilled immigrant. The results hold true when we eliminate states with the greatest or smallest levels of immigration share, states located on the Mexican border, and states with the lowest proportions of immigrants. They are also valid when we exclude undocumented immigrants and those who arrived in the US at a young age. In addition, we find no evidence of an inverted-U shaped relationship as found by Ashraf and Galor (2013) for genetic diversity, or of a negative epidemiological effect *a la* Collier (2013) and Borjas (2015). On the contrary, we find that immigrants from richer countries have a smaller effect on GDP per capita than those from poorer countries; we interpret this as a confirmation that diversity among college-educated immigrants matters more than the economic conditions at origin. Finally, birthplace diversity is negatively correlated with the index of polarization in the immigrant population. If, instead of diversity, a high-skilled polarization index is used, we obtain a highly significant and negative effect on GDP per capita.

To address endogeneity issues, we combine Placebo tests with IV regressions; as far as the latter are concerned, we consider two instrumentation strategies that have been used in the related literature. The first one is a shift-share strategy *a la* Ottaviano and Peri (2006) which includes the predicted diversity indices based on total US immigration stocks by country of origin, and the bilateral state shares observed in 1960. The second strategy consists in instrumenting diversity indices, using the immigration predictions of a pseudo-gravity regression that include interactions between year dummies and the geographic distance between each country of origin and each state of destination (in line with Feyrer (2009) or Alesina *et al.* (2016)). In both cases, diversity among college-educated migrants remains highly significant, while diversity among the less educated is insignificant or weakly significant. In the preferred specification, the coefficient of high-skilled diversity is equal to 0.616. At first glance, this seems important because the average diversity index among college-educated immigrants equals 0.937 in 2010; hence, increasing diversity from zero to 0.937 increases GDP per capita by 58%. However, in 2010, the high-skilled diversity index ranges from 0.797 to 0.976. If all US states had the same level of diversity as the District of Columbia (0.976), the average GDP per capita of the US would be 2.33% larger, the coefficient of variation across states would be 2.37% smaller, and the Theil index would decrease by 3.45%, only. By comparison, if all US states had the same average level of human capital as the District of Columbia, the average GDP per capita of the US would be 8.32% larger, the coefficient of variation across states would be 9.77% smaller, and the Theil index would decrease by 16.06%. Although diversity has non-negligible effects on cross-state disparities, its macroeconomic implications are rather limited.[5] We reach the same conclusion when using the longitudinal dimension of the data. The US-state average level of diversity among college-educated migrants increased by 7 percentage points between 1960 and 2010; this explains a 3.5% increase in macroeconomic performance

[5] The GDP per capita of Hawaii (diversity index of 0.797) would be 11.66% larger if Hawaii had the same diversity index as the District of Columbia; the difference in high-skilled diversity explains about 4.7% of the total income gap between these two states in 2010.

(i.e. only one fiftieth of the total change in the US level of GDP per capita).

The remainder of the paper is organized as follows. Section 2.2 describes our main diversity measures and documents the global trends in cultural diversity in the aftermath of WW2. Section 3.4 describes our empirical strategy. The results are discussed in Section 4.4. Finally, Section 2.5 concludes.

2.2 Diversity in the Aftermath of WW2

Following Ottaviano and Peri (2006), Ager and Brückner (2013), Suedekum *et al.* (2014) and Alesina *et al.* (2016), we consider that the cultural identity of individuals is mainly determined by their country of birth. The rationale is that the competitiveness of modern-day economies is closely linked to the average level of human capital of workers and to the complementarity between their skills. Workers originating from different countries were trained in different school systems and are more likely to bring complementary skills, cognitive abilities and productive traits. In our benchmark model, our key explanatory variable is an index of birthplace diversity (or birthplace fractionalization), which can be computed for each US state and for the high-skilled and low-skilled populations separately. In subsection 2.2.1, we first define various measures of birthplace diversity, establish links between them, and discuss their statistical correlation with the average immigration rate. In subsection 2.2.2, we then document the global US trends in cultural diversity observed in the aftermath of WW2.

2.2.1 The Birthplace Diversity Index

In line with existing studies, we first define a Herfindahl-Hirschmann index of birthplace diversity, $TD^S_{r,t}$, which can be computed for the skill group $S = (L, H, A)$ (L for the low skilled, H for the high skilled, and A for both groups), for each region $r = (1, ..., R)$ and for each year $t = (1, ..., T)$. Our index measures the probability that two randomly-drawn individuals from the type-S population of a particular region originate from two different countries of birth. As shown by Alesina *et al.* (2016) in a cross-country setting, the birthplace diversity index is poorly correlated with genetic or ethnolinguistic fractionalization indices. The index is defined as:

$$TD^S_{r,t} = \sum_{i=1}^{I} k^S_{i,r,t}(1 - k^S_{i,r,t}) = 1 - \sum_{i=1}^{I} (k^S_{i,r,t})^2, \tag{2.1}$$

where $k^S_{i,r,t}$ is the share of individuals of type S, born in country i, and living in region r, in the type-S resident population of the region at year t. Computing the birthplace diversity index requires collecting panel data on the structure of the population by region of destination, by country of origin, and by education level. Our sample includes all US states (including the District of Columbia) between 1960 and 2010 in ten-year intervals, i.e. $r = (1, ..., 51)$ and $t = (1960, ..., 2010)$. Our choice to conduct the analysis at the state level is guided by the availability of long-term data series on macroeconomic performance, and by the comparability with cross-country results. We identify a common set of 195 countries of origin, including the US as a whole.[6] In the Appendix, we conduct the analysis at the level of US Commuting Zones, using wage proxies as dependent variables.[7]

Building on Alesina *et al.* (2016), the additive decomposition of the diversity index allows to distinguish between the *Between* and the *Within* components of the diversity index, $TD^S_{r,t} = BD^S_{r,t} + WD^S_{r,t}$. On the one hand, the *Between* component $BD^S_{r,t}$ measures the probability that a randomly-drawn pair of type-S residents includes a native and an immigrant, irrespective of where the immigrant comes from:[8]

$$BD^S_{r,t} = 2k^S_{r,r,t}(1 - k^S_{r,r,t}).$$

On the other hand, the residual *Within* component $WD^S_{r,t}$ measures the

[6] We disregard heterogeneity between US natives born in different states (e.g. a Texan native is considered identical to a Californian one). See subsection 2.2.2 for a detailed description of the data.

[7] Table A18 describes the results obtained for Commuting Zones and for the 1970-2010 period. Commuting Zones are designed to better capture local labor market conditions. We use the data described in Dorn (2009). They cover all US regions and are fully comparable across periods.

[8] In our specific case, $k^S_{r,r,t}$ represent the share of US natives of type S living in region r at time t.

probability that a randomly-drawn pair of type-S residents includes two immigrants born in two different countries:

$$WD^S_{r,t} = \sum_{i \neq r}^{I} k^S_{i,r,t}(1 - k^S_{i,r,t} - k^S_{r,r,t}).$$

In the US context, the evolution of the birthplace diversity index among residents is almost totally driven by the change in the *Between* component of diversity, $BD^S_{r,t}$, which only depends on the proportion of immigrants. The median share of the *Between* component in total diversity, $BD^A_{r,t}/TD^A_{r,t}$, equals 0.98% and its quartiles are equal to 0.92% and 0.97%. Similar findings are found for the low-skilled and high-skilled populations. Consequently, birthplace diversity in group S is almost perfectly correlated with the region-wide proportion of immigrants.[9] On average, the Pearson correlation between $TD^S_{r,t}$ and the total share of immigrants in the population, $m^S_{r,t} = (1 - k^S_{r,r,t})$, equals 0.99 for all S. It is thus impossible to statistically disentangle the effects of diversity from those of immigration. For this reason and in line with existing works, our empirical specification distinguishes between the size of immigration and the variety of immigrants.

To capture the variety effect, we start from the *Within* component of the diversity index. The *Within* component can be expressed as the product of the square of the immigration rate (the probability that two randomly-drawn individuals are immigrants) by an index of diversity among immigrants, $MD^S_{r,t}$. The latter measures the probability that two randomly-drawn immigrants from region r originate from two different countries of birth. We have:

$$\begin{aligned} WD^S_{r,t} &= (1 - k^S_{r,r,t})^2 MD^S_{r,t} \\ &= (1 - k^S_{r,r,t})^2 \sum_{i \neq r} \widehat{k}^S_{i,r,t}(1 - \widehat{k}^S_{i,r,t}), \end{aligned} \tag{2.2}$$

where $\widehat{k}^S_{i,r,t} = k^S_{i,r,t}/(1 - k^S_{r,r,t})$ is the share of immigrants from origin country i in the total immigrant population of region r. Contrary to the total index of

[9] This is shown in Table A19 in the Appendix, which provides correlations between diversity indices, and between diversity and the immigration rate.

diversity and to its *Between* and *Within* components, the correlation between $MD^S_{r,t}$ and the total immigration rate, $m^S_{r,t}$, is small (on average, -0.19). This allows us to simultaneously include these two variables in the same regression without fearing collinearity problems.

2.2.2 Diversity in the US states

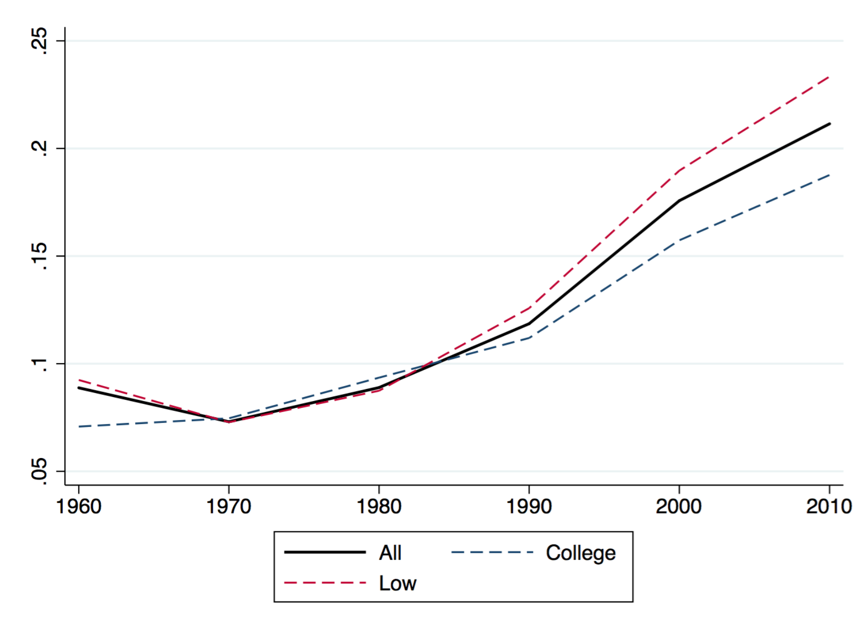
Population data at the state level for the US are available from the Integrated Public Use Microdata Series (IPUMS). IPUMS data are drawn from the federal census of the American Community Surveys. For each census year, they allow characterizing the evolution of the American population by country of birth, age, level of education, and year of arrival in the US, among others. We extracted the data from 1960 to 2010 in ten-year intervals, using the 1% census sample for the years 1960 and 1970, the 5% census sample for the years 1980, 1990 and 2000, and the American Community Survey (ACS-1%) sample for the year 2010. Regarding the origin countries of immigrants, we consider the full set of countries available in 2010, although some of them had no legal existence in the previous census years. Hence, for the years 1960 to 1990, data for the former USSR, former Yugoslavia and former Czechoslovakia are split using the country shares observed in the year 2000. In addition, we treat five pairs of countries as a single entity; this is the case of East and West Germany, Kosovo, Serbia and Montenegro, North and South Korea, North and South Yemen, and Sudan and South Sudan. Finally, we allocate individuals with a non-specified (or an imperfectly specified, respectively) country of birth proportionately to the country shares in the US population (or to the country shares in the US population originating from the reported region, respectively).

In our benchmark regressions, we restrict our micro sample to all individuals aged 16 to 64, who are likely to affect the macroeconomic performance of their state of residence. We distinguish between two skill groups. Individuals with at least one year of college are classified as highly skilled, whereas the rest of the population is considered as low skilled. We define as US natives all individuals born in the US or in US-dependent territories such as American Samoa, Guam, Puerto Rico, the US Virgin Islands and other US possessions.

Other foreign-born individuals are referred to as immigrants.

In alternative regressions, we only consider immigrants who arrived in the US after a certain age, or immigrants who are likely to have a legal status. As for the age-of-entry correction, we sequentially eliminate immigrants who arrived before the age of 5, 6, ... , 25. In order to proxy the number of undocumented immigrants, we follow the "residual methodology" described in Borjas (2016), and use information on the respondents' characteristics (such as citizenship, working sector, occupation, whether they receive public assistance, etc.).

Figure 2.1: Trends in total birthplace diversity in US states, $(TD^S_{r,t})$ 1960-2010

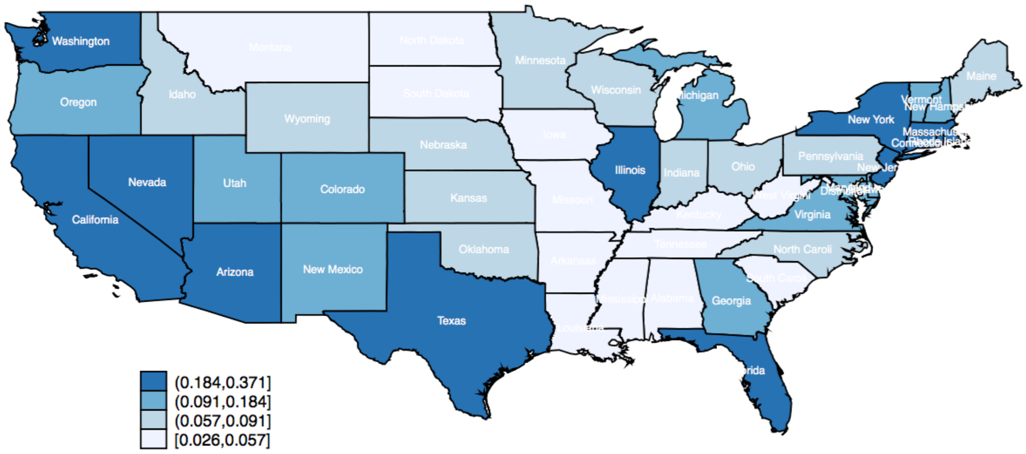


Notes: Diversity among residents is defined as in Eq. (2.1). Source: Authors' elaboration on IPUMS data.

We use IPUMS data to identify the bilateral stocks and shares of international migrants, $k^S_{i,r,t}$, in the population of each state r, by country of origin i and by education level S in the year t. We thus construct comprehensive matrices of "Origin $\times$ State $\times$ Skill" stocks and shares from 1960 to 2010 in ten-year intervals.[10] Missing observations are considered as zeroes, even if a

[10] We distinguish between 195 countries of birth and 50 US states plus the District of

Figure 2.2: Cross-state differences in birthplace diversity among residents ($TD^A_{r,t}$), 1960-2010 average index



Notes: Diversity among residents is defined as in Eq. (2.1). The map presents the average birthplace diversity observed between 1960 and 2010. Alaska and Hawaii are not represented. Source: Authors' elaboration on IPUMS data.

positive number of immigrants is identified for an adjacent year.[11] The evolution of the average index of cultural diversity is described in Figures 2.1 and 2.3 , whereas Figures 2.2 and 2.4 represent differences in the average level of diversity across US states.

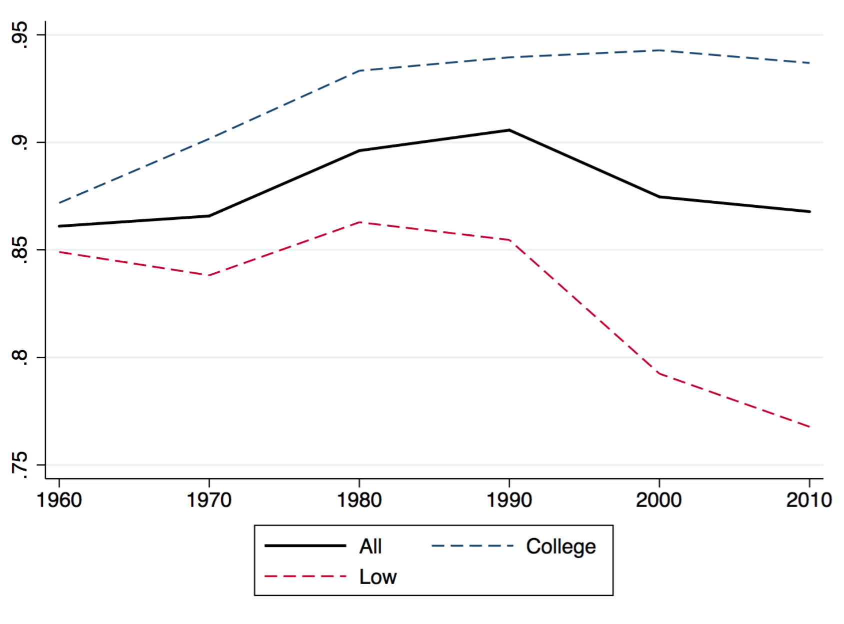
Figure 2.1 describes the evolution of the birthplace diversity index computed for the resident population, $TD^S_{r,t}$ for all S, between 1960 and 2010. Looking at the average of all US states, the birthplace diversity index among residents increased from about 0.09 in 1960 to 0.21 in 2010, reflecting the general rise in immigration to the US. A large portion of this change occurred after 1990. Nevertheless, this average trend conceals important differences between US states and between skill groups. As far as cross-state differences are concerned, the number of immigrants drastically increased in states such

Columbia. Countries and states are listed in Appendix A8. Descriptive statistics by state are provided in Table A9.

[11] The number of zeroes equals 33,145 out of a sample of 59,670 observations (55.5%). The missing values are mostly concentrated in the years 1960 and 1970.

as California (+195%) or New York (+91%); on the contrary, the number of foreign-born individuals remained small and stable in other states such as Montana or Maine. Regarding differences between skill groups, changes in immigration rates were larger for the low skilled than for college graduates, particularly after the year 1980. This is mainly due to the large inflows of low-skilled Mexicans observed during the last three decades, which drastically affected the level of diversity in states located on the West Coast and along the US-Mexican border, as illustrated by Figure 2.2.

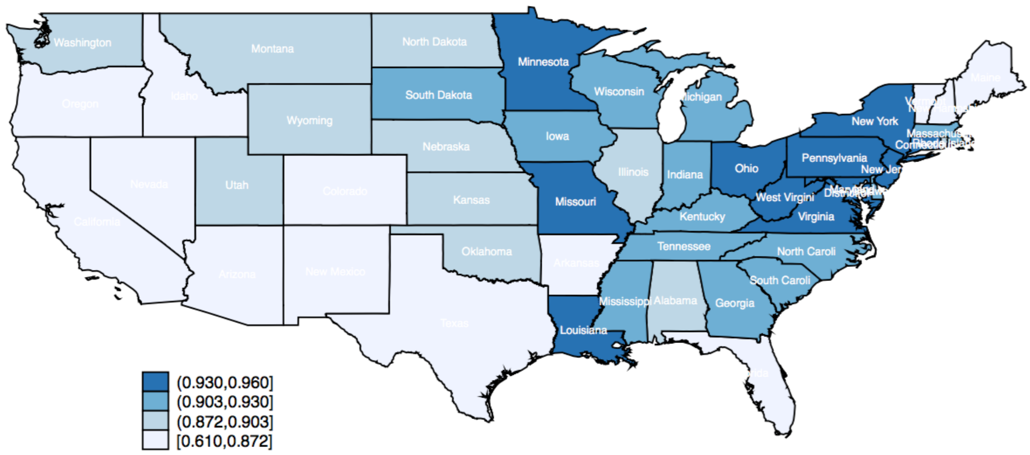
Figure 2.3: Trends in birthplace diversity among immigrants in US states ($MD^S_{r,t}$), 1960-2010



Notes: Diversity among immigrants is defined as in Eq. (2.2). Source: Authors' elaboration on IPUMS data.

Second, Figure 2.3 describes the evolution of the diversity index computed for the immigrant population, $MD^S_{r,t}$ for all S. It shows that on average, the level of diversity in the immigrant population varies across skill groups. Diversity among college-educated immigrants has always been greater than diversity among the less educated. This might be due to the fact that college-educated migrants are less prone to concentrate in regions where large migration networks exist; they consider moving to more (geographically) diversified locations. Differences between skill groups drastically increased after 1960. On the one hand, diversity among high-skilled immigrants increased during the sixties and seventies, possibly due to the the Immigration and Nationality Act of 1965. Changes have been smaller since 1980 despite the Immigration Act of 1990, which allocated 50,000 additional visas (in the form of a lottery) to people from non-typical origin countries. On the other hand, diversity among low-skilled immigrants has fallen since 1980. Again, the latter decline is mainly explained by the large inflows of low-skilled Mexicans. Along the Mexican border and on the West Coast, the probability that two randomly-drawn immigrants were born in two different countries decreased as the share of Mexicans increased. This is also illustrated in Figure 2.4, which reveals important cross-state differences in the long-run average level of diversity among immigrants.

Figure 2.4: Cross-state differences in birthplace diversity among immigrants ($MD^A_{r,t}$), 1960-2010 average index



Notes: Diversity among immigrants is defined as in Eq. (2.2). The map presents the average birthplace diversity observed between 1960 and 2010. Alaska and Hawaii are not represented. Source: Authors' elaboration on IPUMS data.

In sum, the evolution of diversity among immigrants varies across US states and over time. Figure A12 in the Appendix reveals that diversity among immigrants decreased in states located along the US-Mexican border and on the West Coast. A rise in diversity was observed in other states (such as Maine or Vermont). Our panel data analysis takes advantage of these intra-state and inter-state variations to identify a causal effect of diversity on macroeconomic performance.

2.3 Empirical Strategy

Our goal is to identify the effect of multiculturalism on the macroeconomic performance of US states.[12] The level of macroeconomic performance is measured by the log of the Gross Domestic Product (GDP) per capita. In subsection 2.3.1, we present the benchmark specification in which multiculturalism is proxied by the skill-specific indices of birthplace diversity described in Section 2.2. In subsection 2.3.2, we conduct a large set of robustness checks, considering alternative sub-samples, alternative measures of birthplace diversity, and alternative technologies of transmission of cultural shocks. Subsection 2.3.3 explain how we deal with endogeneity issues. We rely on Placebo and IV regressions, combining two instrumentation strategies. Finally, subsection 2.3.4 presents the data sources used to construct our control variables and instruments.

2.3.1 Benchmark Specification

Our benchmark empirical model features the log of GDP per capita as the dependent variable. In line with Ottaviano and Peri (2006), Ager and Brückner (2013), Suedekum *et al.* (2014) and Alesina *et al.* (2016), we use the following specification:

$$log(y_{r,t}) = \beta_1 MD^S_{r,t} + \beta_2 m^S_{r,t} + \boldsymbol{\lambda}' \boldsymbol{X}_{r,t} + \gamma_r + \gamma_t + \varepsilon_{r,t}, \tag{2.3}$$

[12] In the Appendix, a complementary analysis is conducted on the 34 OECD member states, using population data from Özden *et al.* (2011). The first drawback of the database is that it does not report the educational structure of migration stocks. To capture skill-specific effects, we combine it with the 1990-2000 estimates of the bilateral proportion of college graduates provided in Artuc *et al.* (2015). The second drawback is that it relies on imputation techniques to fill the missing bilateral cells. Despite the lower quality of the data, our fixed-effect analysis globally confirms the results obtained for US states.

where $log(y_{r,t})$ is the log of GDP per capita in region r at year t, $MD^S_{r,t}$ is the type-S birthplace diversity among immigrants (proxy for the variety of immigrants), and $m^S_{r,t}$ is the proportion of immigrants in the working-age population of type S. The latter variable captures the other channels through which the level of immigration affects macroeconomic performance (e.g. labor market, fiscal or market-size effects). We opt for a static specification and assume that changes in diversity fully materialize within 10 years. This spares us from dealing with the endogeneity of the lagged dependent, an important issue in dynamic models with a short-panel dimension (Nickel, 1981).[13]

The coefficient β_1 is our coefficient of interest. It captures the effect of diversity on macroeconomic performance. Using skill-specific measures of cultural diversity and immigration, $S = (L, H, A)$, we can identify whether the level and significance of β_1 vary across skill groups. We first estimate Eq. (2.3) using pooled OLS regressions, bearing in mind that such regressions raise a number of econometric issues that might generate inconsistent estimates. The key issue when using pooled OLS regressions is the endogeneity of the main variable of interest, the index of diversity. Endogeneity can be due to a number of reasons. These reasons include the existence of uncontrolled confounding variables causing both dependent and independent variables, the existence of a two-way causal relationship between these variables, or a measurement problem.

To mitigate the possibility of an omitted variable bias, the benchmark model includes a vector $\boldsymbol{X}_{i,t}$ of time-varying covariates. It includes the log of population, the log of the region-wide average educational attainment of the working-age population (as measured by the years of schooling or highest degree completed), and the log of the urbanization rate. In addition, our specification includes a full set of region and year fixed effects, γ_r and γ_t, which allows us to better account for unobserved heterogeneity (including initial conditions in 1960). To solve the reverse causation and measurement problems, we use Placebo tests and two methods of instrumental variables described in

[13]Nevertheless, Tables A25 and A26 in the Appendix provide the results of dynamic GMM regressions with internal or external instruments, and with different lag structures. In these regressions, the lagged dependent is insignificant or weakly significant, which reinforces the credibility of our static benchmark specification. In addition, the effect of diversity is similar to that obtained in the static model.

subsection 2.3.3.

2.3.2 Alternative Specifications

Our benchmark specification Eq. (2.3) assumes linear effects of the level of immigration and of the variety of immigrants on the log of GDP per capita. The literature on multiculturalism suggests that the technology of transmission of cultural shocks can be different.[14]

First, looking at the effect of genetic diversity on economic development, Ashraf and Galor (2013) and Ashraf *et al.* (2015) consider a quadratic specification, which allows them to identify an optimal level of diversity. In our context, cultural diversity may also induce costs and benefits, implying that its effect on macroeconomic performance could be better captured by an inverted-U shape relationship. We thus naturally extend our benchmark specification in sub-section 2.4.2 by adding the square of the birthplace diversity index.

Second, another strand of the literature focuses on migration-induced transfers of norms, and tests for potential epidemiological or contamination effects. Transfers of norms from origin to destination countries have been examined by a limited set of studies.[15] Comparing the economic performance of US counties from 1850 to 2010, Fulford *et al.* (2015) show that the country-of-ancestry distribution of the population matters, and that the estimated effect of ancestry is governed by the sending country's level of economic development, as well as by measures of social capital at origin (such as trust and thrift). Putterman

[14]The birthplace diversity index $MD^S_{r,t}$ does not account for the cultural distance between origin and destination countries. It assumes that all groups are culturally equidistant from each other. Another extension consists therefore in multiplying the probability that two randomly-drawn immigrants were born in two different countries by a measure of cultural distance between these two countries. For the latter, we use the database on genetic distance between countries, constructed by Spolaore and Wacziarg (2009). Genetic distance is based on blood samples and proxies the time since two populations had common ancestors. It is worth noticing that our results are robust to the use of an augmented diversity index and are reported in the Appendix.

[15]More studies focus on emigration-driven contagion effects, i.e. the effects of migrants' destination-country characteristics on outcomes at origin. The most popular study is that of Spilimbergo (2009), which investigates the effect of foreign education on democracy. Beine *et al.* (2013) and Bertoli and Marchetta (2015) use a similar specification to examine the effect of emigration on source-country fertility. Lodigiani and Salomone (2012) find that emigration to countries with greater female participation in parliament increases female participation in the origin country.

and Weil (2010) study the effect of ancestry in a cross-country setting, and find that the ancestry effect is governed by a measure of state centralization in 1500. More recently, debates about the societal implications of diversity have been revived in the migration literature. Collier (2013) and Borjas (2015) emphasize the social and cultural challenges that movements of people may induce. Their reasoning is the following: by importing their "bad" cultural, social and institutional models, migrants may contaminate the set of institutions in their country of adoption, influencing the world distribution of technological capacity. To account for such epidemiological effects, we supplement our benchmark specification with $MY^S_{r,t}$, the weighted average of the log of GDP per capita in the origin countries of type-S immigrants to region r (weights are equal to the bilateral shares of immigrants). The epidemiological term is defined as:

$$MY^S_{r,t} = \sum_{i \neq r}^{I} \widehat{k}^S_{i,r,t} log(y_{i,t}). \tag{2.4}$$

On average, the correlation between this term and the diversity index is small (around -0.17 across US states), so that both variables can be tested jointly. Similarly, the correlation with the immigration rate is rather small (-0.26). Alesina *et al.* (2016) control for such epidemiological terms and find insignificant effects. Compared to them, we consider several variants of Eq. (2.4) in the Appendix, and we also instrument epidemiological terms.

2.3.3 Identification Strategy

Although our benchmark specification includes time-varying covariates and a full range of fixed effects, the positive association between diversity and macroeconomic performance can be driven by reverse causality. As argued by Alesina and La Ferrara (2005), diversity is likely to respond to changes in the economic environment. In particular, economic prosperity and the degree of diversification of production are likely to attract people from different cultural origins. Causation is hard to establish with cross-sectional data. Two methods are used in this paper.

On the one hand, we augment our benchmark specification with natives' migration rates (denoted by $n^S_{r,t}$), and measures of diversity computed for the native population (denoted by $ND^S_{r,t}$). More precisely, we use the IPUMS data to identify the state of birth and the state of residence of each American citizen, and we compute internal migration rates and indices of diversity by state of birth for both skill groups. The latter index measures the probability that two randomly-drawn Americans from the type-S population of a particular state originate from two different states of birth. If diversity responds to economic prosperity, we expect a positive correlation between $ND^S_{r,t}$ and GDP per capita. On the other hand, we use a two-stage least-square estimation method. We compare the results obtained under alternative sets of instruments, and show that our IV results are robust to the instrumentation strategy. We consider two different sets of instruments that have been used in the existing literature.

Our first IV strategy is a shift-share strategy *a la* Ottaviano and Peri (2006) or Ager and Brückner (2013). The set of instruments includes an index of remoteness, as well as predicted diversity indices based on total US immigration stocks by country of origin, and bilateral shares observed in 1960. Following the shift-share methodology, we predict the skill-specific bilateral migration stocks for each state using the residence shares of natives and immigrants observed in 1960. Then, we use these shares to allocate the new immigrants by state of destination. The predicted stock of migrants at time t is:

$$\widehat{Stock^S}_{i,r,t} = Stock^S_{i,r,1960} + \phi^S_{i,r}(Stock^S_{i,t} - Stock^S_{i,1960}), \tag{2.5}$$

where $Stock^S_{i,r,t}$ is the type-S stock of immigrants from country i residing in region r at year t. The term $\phi^S_{i,r}$ is the time-invariant share that we use to allocate the variation in the bilateral migration stocks observed between the years 1960 and t. More precisely, we allocate changes in bilateral migration stocks using the 1960 skill-specific shares of US natives and immigrants from the same origin country. These shares capture both origin- and skill-specific network effects, and the concentration of type-S workers in 1960. We have:

$$\phi^S_{i,r} = \frac{Nat^S_{r,1960} + Stock^S_{i,r,1960}}{\sum_r (Nat^S_{r,1960} + Stock^S_{i,r,1960})}, \tag{2.6}$$

where $Nat^S_{r,1960}$ is the number of US natives residing in region r at year 1960. Using the predicted stock of migrants (who are less likely to be affected by the economic performance of each state), we compute the predicted diversity indices.

In line with Feyrer (2009) or Alesina *et al.* (2016), our second IV strategy consists in instrumenting diversity indices using the predicted migration stocks obtained from a "zero-stage", pseudo-gravity regression. The latter regression includes interactions between year dummies and the geographic distance between each country of origin and each US state. In line with the shift-share strategy, the identification thus comes from the time-varying effect of geographic distance on migration, reflecting gradual changes in transportation and communication costs. The pseudo-gravity model is written:

$$\log(Stock_{i,r,t}) = \beta_t \log(Dist_{i,r}) + Bord_{i,r} + Lang_{i,r} + \gamma_r + \gamma_i + \gamma_t + \varepsilon_{i,r,t}, \quad (2.7)$$

where $Bord_{i,r}$ is a dummy equal to one if country i and region r share a common border, $Lang_{i,r}$ is a dummy equal to one if at least 9% of the populations of i and r speak a common language, γ_r, γ_i, and γ_t are the destination, origin and year fixed effects. In the pseudo-gravity stage, the high prevalence of zero values in bilateral migration stocks gives rise to econometric concerns about possible inconsistent OLS estimates. To address this problem, we use the Poisson regression by pseudo-maximum likelihood (see (Santos Silva and Tenreyro, 2006)). Standard errors are robust and clustered by country-state pairs.

Although commonly used in the literature, each of these IV strategies has some drawbacks. The augmented shift-share and internal methods are imperfect if potential regressors exhibit strong persistence. In addition, the relative geography variables used in the strategy *a la* Feyrer (2009) can affect macroeconomic performance through other channels such as trade, foreign direct investments or technology diffusion (not measurable at the state level for the 1960-2010 period). Nevertheless, we can reasonably support a careful causal interpretation of our results if these strategies yield consistent and converging results.

Table 2.1: Summary statistics 1960-2010

	Mean	Std.D	Min	Max
$TD^A_{r,t}$	0.126	0.105	0.006	0.548
$MD^A_{r,t}$	0.879	0.099	0.342	0.974
$m^A_{r,t}$	0.068	0.061	0.003	0.347
$TD^H_{r,t}$	0.116	0.087	0.007	0.478
$MD^H_{r,t}$	0.921	0.054	0.610	0.976
$m^H_{r,t}$	0.061	0.049	0.003	0.281
$TD^L_{r,t}$	0.134	0.121	0.006	0.592
$MD^L_{r,t}$	0.827	0.141	0.293	0.967
$m^L_{r,t}$	0.074	0.073	0.003	0.417
$log(y_{r,t})$	9.534	1.018	7.587	12.058
$log(Pop_{r,t})$	14.390	1.068	11.831	17.042
$log(Urb_{r,t})$	4.201	0.245	3.472	4.605
$log(Hum_{r,t})$	1.806	0.156	1.360	2.072

Source: Authors' elaboration on IPUMS-US data.

2.3.4 Data Sources

The sources of our migration data were described in Section 2.2. In this subsection, we describe the data sources used to construct our dependent variables, the set of control variables, and the set of instruments. Table 2.1 summarizes the descriptive statistics of our main variables. More details on our data sources and variable definitions are available in Table A7 in the Appendix. The data for GDP ($y_{r,t}$) are provided by the Bureau of Economic Analysis for US states. The population data by age are taken from the IPUMS database. We consider the population aged 15 to 64 ($Pop_{r,t}$) in the regressions. The US Bureau of Census also provides the data on urbanization rates for US states ($Urb_{r,t}$); the urbanization rate measures the percentage of the population living in urbanized areas, and urban clusters are defined in terms of population size and density. As for human capital ($Hum_{r,t}$), we compute the average educational attainment of the working-age population using the IPUMS database.

As far as the set of instruments is concerned, the data on geographic distance between origin countries and US states are computed using the latitude and the longitude of the capital city of each US state and each country. Such data are available from the Infoplease and Realestate3d websites which have allowed us to compute a bilateral matrix of great-circle distances between US state capital cities and countries.[16]

2.4 Results

Our empirical analysis follows the structure explained in Section 3.4. In subsection 2.4.1, we investigate the effect of birthplace diversity among immigrants using pooled OLS regressions; we produce separate results for the three skill groups of immigrants. Then, we test for the existence of epidemiological effects, and we control for unobserved heterogeneity, including a full set of state and year fixed effects (FE). In subsection 2.4.2, we show that the FE estimates are robust to sub-samples with the exclusion of states with the greatest or smallest immigration rates, or states sharing a common border with Mexico. We also show that our results are stable when controlling for the share of the ten largest groups of immigrants, when considering alternative education categories, or when using alternative diversity indices. In other robustness checks, we take into account the legal status and the age of entry of migrants, and we test for possible non-linear effects of birthplace diversity. Finally, in subsection 2.4.3, we address endogeneity issues using Placebo and IV regressions; the latter rely on two instrumentation strategies frequently used in the existing literature, i.e a shift-share strategy *a la* Ottaviano and Peri (2006) and a gravity-like strategy *a la* Feyrer (2009).

2.4.1 Pooled OLS and FE Regressions

Table 2.2 describes the pooled OLS and FE estimates. We produce separate results for the three skill groups, $S = (A, L, H)$, under the same set of control variables, including the skill-specific immigration rate, $m^S_{r,t}$, the log of population, $log(Pop_{r,t})$, the log of urbanization, $log(Urb_{r,t})$, and the log of the average educational attainment of the working-age population, $log(Hum_{r,t})$. In all cases, our standard errors are clustered at the state level in order to correct for heteroskedasticity and serial correlation.

[16] See `http://www.infoplease.com/ipa/A0001796.html` and `http://www.`

Table 2.2: Pooled OLS and FE regressions
Results by skill group (Dep= $log(y_{r,t})$)

	(1) OLS $S=A$	(2) FE $S=A$	(3) OLS $S=L$	(4) FE $S=L$	(5) FE $S=L$	(6) OLS $S=H$	(7) FE $S=H$	(8) FE $S=H$
$MD^S_{r,t}$	0.416	0.318***	0.019	0.141	0.104	2.719***	0.616***	0.531***
	(0.329)	(0.114)	(0.184)	(0.086)	(0.085)	(0.719)	(0.160)	(0.159)
$m^S_{r,t}$	2.632***	0.582*	1.901***	0.481*	0.412	4.383***	0.614*	0.388
	(0.615)	(0.341)	(0.485)	(0.282)	(0.283)	(1.018)	(0.315)	(0.366)
$MY^S_{r,t}$					-0.104**			-0.133*
					(0.042)			(0.069)
$log(Pop_{r,t})$	0.070	-0.172**	0.079*	-0.166**	-0.146*	0.011	-0.155**	-0.080
	(0.047)	(0.079)	(0.047)	(0.081)	(0.082)	(0.044)	(0.075)	(0.065)
$log(Urb_{r,t})$	-0.407*	0.385**	-0.367	0.329**	0.312*	-0.563**	0.285**	0.156
	(0.238)	(0.156)	(0.254)	(0.163)	(0.173)	(0.229)	(0.135)	(0.138)
$log(Hum_{r,t})$	5.752***	0.695***	5.817***	0.807***	0.802***	5.288***	0.759***	1.007***
	(0.157)	(0.197)	(0.147)	(0.205)	(0.196)	(0.182)	(0.197)	(0.299)
Constant	-0.697	7.529***	-0.728	7.662***	8.379***	-0.584	7.348***	7.492***
	(0.890)	(1.254)	(0.914)	(1.263)	(1.317)	(0.890)	(1.262)	(1.273)
Observations	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51
R-squared	0.879	0.993	0.878	0.993	0.993	0.889	0.993	0.993
Time fixed effects	No	Yes	No	Yes	Yes	No	Yes	Yes
States fixed effects	No	Yes	No	Yes	Yes	No	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (2.3). Pooled OLS results are provided in col. 1, 3 and 6; FE results are provided in col. 2, 4, 5, 7 and 8. Results for all immigrants are provided in col. 1 and 2; results for low-skilled immigrants are provided in col. 3, 4 and 5; results for college-educated immigrants are provided in col. 6, 7 and 8. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$). We supplement our benchmark specification in col. 5 and 8 with the epidemiological effect ($MY^S_{r,t}$).

The pooled OLS estimates are reported in col. 1, 3 and 6. We find that the effect of birthplace diversity on GDP per capita is skill-specific. Insignificant effects are obtained when diversity is computed using the low-skilled or the total immigrant populations.[17] On the contrary, the association between GDP per

`realestate3d.com/gps/latlong.htm` (accessed on July 4, 2016).

[17]When computing the diversity index on the total immigrant population, the effect is significant at the 1% level in the FE regression. This is because high-skilled diversity influences this index.

capita and birthplace diversity among college-educated immigrants is positive and significant at the 1% level. The coefficient is large, implying that a 10% increase in high-skilled diversity is associated with a 27.2% increase in GDP per capita.[18]

In col. 2, 4 and 7, we present the benchmark results, introducing state and year fixed effects in order to mitigate the omitted variable bias. The state fixed effects account for all time-invariant state characteristics that could jointly affect productivity and diversity; the year fixed effects account for time-varying sources of change in GDP per capita that are common to all US states. In the FE regressions, the R-squared is above 0.99. The effect of diversity remains highly significant for college-educated immigrants, and remains insignificant for the less educated. Interestingly, the inclusion of fixed effects leads to a drop in our estimated diversity coefficient. The coefficient of high-skilled diversity is divided by four compared to the pooled OLS regression. This demonstrates that accounting for unobserved heterogeneity is crucial when addressing such an issue. As for our control variables, human capital and urbanization rates are significantly and positively associated with GDP per capita. On the contrary, the correlation between GDP and population size is negative. More interestingly, immigration rates are always positively associated with GDP per capita, and the correlation is always greater for college graduates.[19]

In sum, we find that diversity is positively associated with the level of GDP per capita, but only when diversity is computed on workers performing

[18]One could be concerned that some of our controls are endogenous inducing a bias in the coefficient of diversity. We show in Table A15 in the Appendix that our results still hold when removing our controls variables.

[19]Our results hold when using a measure of cultural polarization. In Ager and Brückner (2013), Montalvo and Reynal-Querol (2003) and Montalvo and Reynal-Querol (2005), the index of polarization captures how far the distribution of a population is from the bimodal distribution. It is defined as: $TP^S_{r,t} = 1 - \sum_{i=1}^{I} \left((0.5 - k^S_{i,r,t})/0.5\right)^2 k^S_{i,r,t}$. The rationale is that a more polarized population can be associated with increased social conflict and a reduction in the quality and quantity of public good provision. Applied to the immigrant population (i.e. using $\widehat{k}^S_{i,r,t}$ instead of $k^S_{i,r,t}$ in the previous equation, the index $MP^S_{r,t}$ is maximized when there are two groups of immigrants which are of equal size (i.e. 50%). For US states, the polarization index exhibits a correlation of -0.89 with the fractionalization index, drammatically high in comparison to Ager and Brückner (2013) due to the high level of diversity in our sample in comparison to their data. Hence, including these two variables in the same regression is risky. The results obtained when using the polarization index are reported in Table A14 in the Appendix.

complex or skill-intensive tasks. On the contrary, diversity among less educated immigrants does not have a significant effect on macroeconomic performance. According to our fixed-effect estimates, a 10% increase in high-skilled diversity (i.e. in the probability that two randomly-drawn, college-educated immigrants originate from two different countries of birth) is now associated with a 6.2% increase in GDP per capita. Expressed differently, a one-standard-deviation increase in high-skilled diversity is associated with a 3.2% increase in GDP per capita. This implies that, if all US states had the same level of diversity as the most diverse state in 2010, i.e. the District of Columbia (0.976), the average GDP per capita of the US would be 2.3% larger, the coefficient of variation across states would be 2.4% smaller, and the Theil index would decrease by 3.5%. By comparison, if all US states had the same average level of human capital as the District of Columbia, the average GDP per capita of the US would be 8.3% larger, the coefficient of variation across states would be 9.8% smaller, the Theil index would decrease by 16.1% and the GDP per capita of Hawaii, the least diverse state in 2010 (0.797), would be 11.7% larger. In addition, the US-state average level of diversity among college-educated migrants increased by 7 percentage points between 1960 and 2010; this explains a 3.5% increase in macroeconomic performance (i.e. only one fiftieth of the total change in the US level of GDP per capita). Although diversity has significant effects on cross-state disparities, its macroeconomic implications are rather limited.

Finally, we supplement the benchmark model with epidemiological effects *a la* Collier (2013) and Borjas (2015) in col. 5 and 8. Interpreting the coefficient of the epidemiological term is not straightforward. On the one hand, if immigrants originating from poor countries contaminate the total factor productivity or the quality of institutions at destination, we should find a positive and significant relationship between our epidemiological term ($MY^S_{r,t}$) and macroeconomic performance. On the other hand, if attracting immigrants from economically or culturally distant countries generates more complementarities in skills and ideas than immigrants from richer countries, we should find a negative and significant relationship. Moreover, reverse causality is a serious source of concern as macroeconomic performance affects the attractiveness of states and the variety of their immigrant population. Our database reveals that richer states attract more people, including immigrants from poorer countries. This selection issue pushes the correlation between GDP per capita and the epidemiological term downwards.[20] This reverse causality issue will be addressed in subsection 2.4.3.[21] As far as low-skilled immigrants are concerned, controlling for epidemiological effects in col. 5 , we find a negative and significant coefficient (at the 5% level). Although we suspect that the negative relationship between the epidemiological term and US state level of GDP per capita can be driven by reverse causality, the negative sign suggests that low-skilled immigrants from richer countries generate fewer complementarities with US natives than immigrants from poorer countries, and/or that greater economic growth in a state attracts more immigrants from poorer countries. As far as high-skilled immigrants are concerned, we find no clear evidence of contamination effects driven by high-skilled immigration in col. 8. The epidemiological effect is insignificant, whereas the coefficient of birthplace diversity is hardly affected. Overall, we find no evidence of a significant contamination mechanism.[22]

2.4.2 Robustness checks

This subsection investigates the robustness of our previous results. Tables 2.3 and 2.4 summarize the results for high-skilled and low-skilled immigrants, respectively. These two tables only report the main results for all the robustness checks that have been done. All models include the full vector of controls (not shown) with the log of population ($log(Pop_{r,t})$), the log of urbanization

[20]Figure A7 in the Appendix confirms this presumption. When we keep the levels of GDP per capita constant for all origin countries (at their 1960-2010 average), we observe that the US state level of GDP per capita is negatively correlated with the epidemiological term.

[21]We also consider alternative specifications for the epidemiological term in the Appendix Table A21. We first compute $MY^S_{r,t}$ by keeping the immigration shares ($\widehat{k}^S_{i,r,t}$) constant, at their 1960-2010 average levels. Then, we keep the levels of GDP per capita at origin ($log(y_{i,t})$) constant, at their 1960-2010 average level. Finally, we combine annual data on GDP per capita at origin with individual data on the year of arrival in the US; each immigration share is multiplied by the average level of GDP per capita prevailing in the year of immigration to the US. The latter specification allows us to capture the norms and values that immigrants bring with them when they migrate. Due to data limitations, this variable cannot be computed for the year 1960.

[22]We obtain the same conclusion when the epidemiological term in Eq. (2.4) is based on democracy levels at origin, instead of GDP per capita. We use the Polity2 index of democracy. These unreported results are available upon request.

($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$) as well as time and state fixed effects. Complete tables are provided in Tables A10 to A17 in the Appendix .

Table 2.3: Robustness of FE regressions for high-skilled diversity (Dep= $log(y_{r,t})$)

		Sub-Samples (A10)				
	(1) Full-Sample	(2) 1970-2000	(3) No Top5	(4) No Bot5	(5) No Mex	(6) No Q1
$MD^H_{r,t}$	0.616***	0.870***	0.725***	0.672***	0.630***	0.596**
	(0.160)	(0.321)	(0.174)	(0.170)	(0.170)	(0.288)
$m^H_{r,t}$	0.614*	1.140**	1.317**	0.613*	0.541	0.765**
	(0.315)	(0.459)	(0.529)	(0.323)	(0.397)	(0.365)
Observations	306	204	276	276	282	228
Nb. states	51	51	46	46	47	38
R-squared	0.993	0.990	0.993	0.993	0.993	0.995
	10 Largest (A12)	Quadratic (A14)	Educ. levels (A16)		Legal Status (A17)	
	(7)	(8)	(9) Ph.D	(10) Tertiary	(11) Docum.	(12) Undoc.
$MD^H_{r,t}$	0.617***	-0.131	0.262**	0.369***	1.009**	-0.153
	(0.169)	(1.954)	(0.103)	(0.136)	(0.473)	(0.127)
$m^H_{r,t}$	0.726*	0.622*	0.256	0.372	0.959*	4.426**
	(0.365)	(0.314)	(0.266)	(0.298)	(0.535)	(2.140)
$(MD^H_{r,t})^2$		0.453				
		(1.202)				
Observations	306	306	306	306	204	204
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.979	0.979

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. A-indexed numbers in parentheses refer to full Tables provided in the Appendix. All models include the full vector of controls (not shown) with the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$) as well as time and states fixed effects.

Robustness by subsample. – In Tables 2.3 and 2.4, the benchmark results of Table 2.2 are reported in col. 1. In col. 2, we limit our sample to the 1970-2000 period, eliminating possible sources of variation prior to the 1965 amendments to the Immigration and Nationality Act, as well as variations driven by the recent evolution of diversity.[23] Then, in col. 3 and 4, we examine whether the impact of diversity is driven by the size of the immigrant population: we drop the five US states with the greatest or the smallest immigration rates in 2010, respectively.[24] In col. 5, we investigate whether our results are driven by the Mexican diaspora, which represented 30% of the whole immigrant population of the US in 2010. We drop the states located on the US-Mexican border, which host 62% of all Mexican immigrants to the US.[25] Remember that these states have experienced a drastic decrease in their diversity index (-40% in low-skilled diversity between 1960 and 2010), which is totally due to the rising inflows from Mexico. Finally, in col. 6, we exclude the states in the first quartile (i.e. below Q1) of the 2010 distribution by immigrant population size. Overall, we show that our FE results are robust to sample selection. In Table 2.3, the coefficient of high-skilled diversity is always positive, significant, and of the same order of magnitude as the benchmark estimates in col. 1. The positive impact becomes even larger when reducing the time span (0.87) or after excluding the states with the highest immigration rates (0.73). This suggests that high-skilled diversity could generate non-linear effects on macroeconomic performance (e.g. a decreasing marginal impact); we will explore this hypothesis in col. 8. As for Table 2.4, it shows that low-skilled diversity is insignificant in all specifications but one. It only becomes significant in col. 3, when the most diverse states are excluded, but only at the 5% level.

Controlling for large groups. – We now investigate whether the effect of birthplace diversity does is not driven by the presence of large diasporas characterized by specific productivity levels (this generalizes what we did when excluding states located on the US-Mexican border). To do so, we control for the state-specific shares that the ten largest origin countries in the US immigrant population. In col. 7 of the two tables, we only report the coefficient for

[23] Remember that Figure 2.3 shows that the average high-skilled diversity index slightly decreased between 2000 and 2010.

[24] The states with the greatest immigration rates are California, New York, Hawaii, New Jersey, and Florida. The states with the smallest rates are West Virginia, Mississippi, Kentucky, South Dakota, and Alabama.

[25] These include California, Texas, New Mexico, and Arizona.

Table 2.4: Robustness of FE regressions for low-skilled diversity (Dep= $log(y_{r,t})$)

		Sub-Samples (A11)				
	(1)	(2)	(3)	(4)	(5)	(6)
	Full-Sample	1970-2000	No Top5	No Bot5	No Mex	No Q1
$MD^H_{r,t}$	0.141	0.130	0.228**	0.128	0.109	0.015
	(0.086)	(0.097)	(0.093)	(0.092)	(0.091)	(0.115)
$m^H_{r,t}$	0.481*	0.691**	0.938**	0.448	0.474	0.647**
	(0.282)	(0.272)	(0.418)	(0.288)	(0.392)	(0.276)
Observations	306	204	276	276	282	228
Nb. states	51	51	46	46	47	38
R-squared	0.993	0.989	0.993	0.993	0.993	0.993

	10 Largest (A13)	Quadratic (A14)	Educ. levels (A16)			Legal Status (A17)	
	(7)	(8)	(9)	(10)	(11)	(12)	(13)
			No School	Primary	Secondary	Docum.	Undoc.
$MD^H_{r,t}$	0.276**	0.705**	0.038	0.033	0.120	0.043	-0.017
	(0.104)	(0.293)	(0.032)	(0.061)	(0.101)	(0.112)	(0.045)
$m^H_{r,t}$	0.058	0.504*	0.152	0.165	0.432	1.107**	3.482*
	(0.274)	(0.281)	(0.100)	(0.101)	(0.301)	(0.424)	(1.795)
$(MD^L_{r,t})^2$		-0.391*					
		(0.218)					
Observations	306	306	306	306	306	204	204
Nb. states	51	51	51	51	51	51	51
R-squared	0.994	0.993	0.993	0.993	0.993	0.979	0.979

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. A-indexed numbers in parentheses refer to full Tables provided in the Appendix. All models include the full vector of controls (not shown) with the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$) as well as time and states fixed effects.

diversity. As far as high-skilled migrants are concerned, controlling for the size of the largest immigrant groups neither affects the significance nor the magnitude of our coefficients of interest. As for low-skilled diversity, the diversity coefficient becomes significant but its magnitude is small.

Quadratic specification. – In col. 8, we supplement our benchmark specification with the squared index of birthplace diversity. If an optimal level of diversity exists, we should find a positive coefficient for the linear term, and a negative coefficient for the squared term. As far as high-skilled immigrants are concerned, we find no evidence of a quadratic effect of birthplace diversity. The coefficient for the squared index of diversity is insignificant in Table 2.3. Hence, this regression rejects the existence of an optimal level of diversity among college-educated immigrants. As far as low-skilled immigrants are concerned, the coefficient for the quadratic term is negative and significant, but only at the 10% level in Table 2.4, while the linear term is significant at the 5% level. We cannot reject the possibility of an inverted-U-shaped relationship, with an optimal level of diversity equal to $MD^L_{r,t} = 0.90$, but the interval of confidence of the quadratic effect is large.[26]

Robustness by skill group. – One might be concerned that the positive effect of high-skilled diversity is driven by the presence of immigrants at the very top of the skill distribution. Similarly, it can be suspected that the insignificant effect of low-skilled diversity is due to the the prevalence of immigrants with very low levels of education. We investigate these issues in col. 9 and 10 of Table 2.3 and in col. 9 to 11 in Table 2.4. As far as high-skilled diversity is concerned, we find insignificant differences when computing diversity on PhD graduates, or on other college-educated immigrants. As for low-skilled diversity, the effect remains insignificant when computing the diversity index on the immigrant populations with no schooling, primary education or secondary education.

Robustness by legal status. – We also investigate the role of undocumented migration in governing the skill-specific effects of diversity. The US census counts every person regardless of immigration status. Hence, undocumented immigrants influence our diversity index. This can be a source of concern as undocumented migrants are likely to be less educated than the legal ones and to contribute differently to GDP, either because their productive activities are not recorded in the official GDP or because they are employed in jobs/sectors where skill complementarities are smaller. This could explain why the effect of low-skilled diversity is insignificant in most of our regressions. To explore this hypothesis, we use the "residual methodology" proposed by Borjas (2016)

[26]Note that $MD^L_{r,t} = 0.90$ corresponds to the median of the distribution. The US states with a low-skilled diversity index around the optimal level in 2010 are Rhode Island (0.891) and Michigan (0.903).

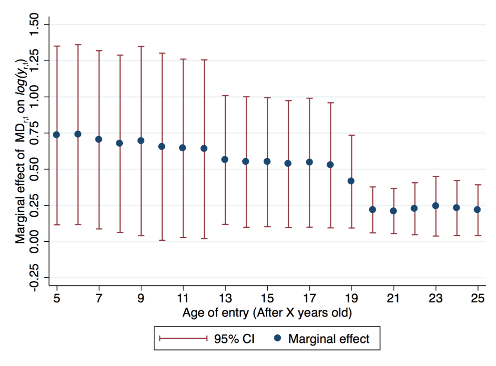
to identify the number of legal and undocumented immigrants by skill group. It consists in using individual characteristics to proxy the legal status of US immigrants. In this work, we use five characteristics (citizenship, employment industry, occupation, whether the individual receives any assistance, and the spouse's legal status) and, due to data availability, we apply the residual methodology to the census years 1980 to 2010. We obtain similar results as in Borjas (2016). For the year 2010, our estimated proportions of undocumented immigrants are equal to 23% in California, 7% in New York, and 15% in Texas; for the same states in the year 2012, Borjas (2016) also obtains 23%, 7%, and 15%. Moreover, the observable characteristics of our undocumented population are also similar. We identify 50% of males and 36% of college graduates; Borjas (2016) obtains 55% and 40%, respectively. As a robustness check, we thus compute the diversity indices on the legal and undocumented immigrant populations, and include them separately in our FE regressions. Col. 11 and 12 in Table 2.3 and Col. 12 and 13 in Table 2.4 give the results for the two skill groups. As far as high-skilled immigrants are concerned, distinguishing between legal and undocumented immigrants yields different effects. Diversity among undocumented immigrants has no significant effect, while diversity among legal immigrants has a positive and significant effect at the five percent level. On the contrary, controlling for the legal status of low-skilled immigrants does not modify our conclusions. It confirms that the insignificant effect of low-skilled diversity cannot be attributed to the greater proportion of undocumented migrants in this group (on average, 17% for the US in 2010).

Robustness by age of entry. – The diversity indices used in our benchmark regressions are computed for the total population of working-age immigrants, whatever their age of entry in the US. As birthplace diversity conceivably reflects complementarities between individuals trained in different countries, it can be argued that immigrants who arrived in the US at different ages generate different levels of complementarity in skills and ideas with the native workforce. However, the role of the age of entry is unclear. On the one hand, immigrants with a longer foreign education are likely to bring more complementarities. On the other hand, immigrants who were partly educated in the

US may have more transferable skills and a greater potential to interact with natives. To investigate this issue, we compute the diversity index using various samples of immigrants, and we include these alternative indices in Eq. (2.3). More precisely, we exclude from the immigrant population the individuals who arrived in the US before a given age threshold, which ranges from 5 to 25 in one-year intervals. For each skill group, Figures 2.5 and 2.6 report the marginal effect of diversity and its confidence interval as a function of the age-of-entry threshold.[27] As information on age of entry is not available in the 1960 census, our sample covers the 1970-2010 period. For this time span, the coefficients of the benchmark FE regressions (without controlling for age of entry) are equal to 0.835 for high-skilled diversity (significant at the 1% level), and to 0.088 for low-skilled diversity (insignificant). Whatever the age-of-entry threshold, the effect of low-skilled diversity is insignificant. Nevertheless, the age of entry matters for college graduates. Although the coefficient of high-skilled diversity is always positive and significant, the largest effects are obtained when the immigrant population includes individuals who arrived before age 20. Considering three age thresholds (12, 18, and 22), Alesina *et al.* (2016) show that the positive effect of birthplace diversity slightly decreases when eliminating children immigrants, but always remains large and significant. Conversely, our results suggest that the greatest levels of complementarity are obtained when immigrants acquired part of their secondary education abroad and their college education in the US.

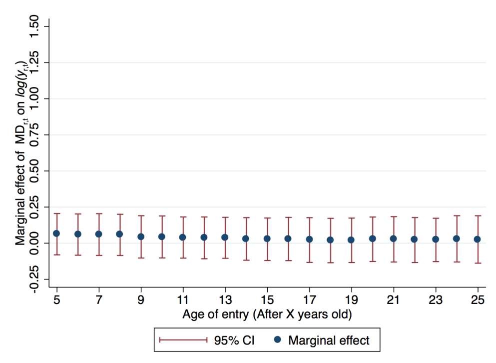
[27] Comprehensive regression results are provided in Table A24 in the Appendix.

Figure 2.5: Marginal effect of $MD^H_{r,t}$ on $log(y_{r,t})$
Results for different age-of-entry thresholds (1970-2010), high-skilled



Source: Authors' elaboration on IPUMS data. Notes: The graph reports the marginal effect of $MD^H_{r,t}$ on $log(y_{r,t})$ when the immigrant population is restricted to individuals who arrived in the US after age X. Marginal effects are obtained using our main specification Eq. (2.3) which includes state and year fixed effects, as well as the immigration rate ($m^H_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

Figure 2.6: Marginal effect of $MD^L_{r,t}$ on $log(y_{r,t})$
Results for different age-of-entry thresholds (1970-2010), low-skilled



Source: Authors' elaboration on IPUMS data. Notes: The graph reports the marginal effect of $MD^L_{r,t}$ on $log(y_{r,t})$ when the immigrant population is restricted to individuals who arrived in the US after age X. Marginal effects are obtained using our main specification Eq. (2.3) which includes state and year fixed effects, as well as the immigration rate ($m^L_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

2.4.3 Dealing with endogeneity

In this section, we investigate the likelihood that reverse causality drives our results. We use Placebo and IV regressions to deal with the endogeneity of birthplace diversity, the immigration rate and the epidemiological term.

Table 2.5: $MD_{r,t}$ v.s diversity among "native immigrants" $ND_{r,t}$
Results by skill group (Dep= $log(y_{r,t})$)

	(1) $S=H$	(2) $S=H$	(3) $S=H$	(4) $S=L$	(5) $S=L$	(6) $S=L$
$MD^S_{r,t}$	0.616***		0.432**	0.141		0.156*
	(0.160)		(0.168)	(0.086)		(0.082)
$m^S_{r,t}$	0.614*		1.008***	0.481*		0.642**
	(0.315)		(0.347)	(0.282)		(0.278)
$ND^S_{r,t}$		0.968	1.183*		0.456	0.513
		(0.654)	(0.662)		(0.570)	(0.563)
$n^S_{r,t}$		0.376**	0.428**		0.059	0.218
		(0.167)	(0.203)		(0.242)	(0.244)
$log(Pop_{r,t})$	-0.155**	-0.135*	-0.176**	-0.166**	-0.120*	-0.169**
	(0.075)	(0.068)	(0.073)	(0.081)	(0.067)	(0.073)
$log(Urb_{r,t})$	0.285**	0.294*	0.316**	0.329**	0.266	0.304*
	(0.135)	(0.159)	(0.149)	(0.163)	(0.164)	(0.172)
$log(Hum_{r,t})$	0.759***	0.557**	0.477**	0.807***	0.677**	0.505*
	(0.197)	(0.213)	(0.217)	(0.205)	(0.271)	(0.269)
Constant	7.348***	6.829***	6.810***	7.662***	7.193***	7.712***
	(1.262)	(1.450)	(1.443)	(1.263)	(1.048)	(1.088)
Observations	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Source: Authors' elaboration on IPUMS-US data. $ND^S_{r,t}$ is computed as the diversity among natives born in a different state than the r state where they reside. $n_{r,t}$ is the immigration rate in the state r where immigrants are natives born in a different state than r.

Placebo regressions. – If diversity increases with economic prosperity, we expect a positive correlation between birthplace diversity among American workers and GDP per capita, as explained in Section 3.4. Table 2.5 reports the results of our Placebo tests. We augment the benchmark model with two additional control variables, namely the natives' migration rates ($n^S_{r,t}$) and the measures of diversity computed for the native population ($ND^S_{r,t}$). It comes out that internal immigration rates are positively correlated with GDP per capita. However, the native diversity index is insignificant (or weakly significant in col. 3). Although these Placebo tests do not necessarily imply that diversity among foreign immigrants is not affected by macroeconomic performance, they mitigate the risk of a strong reverse causation relationship.

IV regressions. – Table 2.6 the results of our 2SLS regressions. In col. 1, 2, 6 and 7, we first only instrument our main variable of interest, $MD^S_{r,t}$, and use the two IV strategies detailed in subsection 2.3.3. The first one is a shift-share strategy, which uses the predicted diversity index based on the 1960 geographic structure of each bilateral diaspora. The second one is the gravity-like strategy *a la* Feyrer (2009). First-stage estimates are provided in Tables A22 and A23 in the Appendix. Then, in the remaining columns, we deal with the endogeneity of two other important regressors, the immigration rate ($m^S_{r,t}$) and the epidemiological term ($MY^S_{r,t}$). To do so, we use the gravity-like strategy *a la* Feyrer (2009) only.

Table 2.6 confirms our previous findings for diversity among high-skilled migrants when only $MD^H_{r,t}$ is instrumented in col. 1 and 2. The effect of $MD^H_{r,t}$ is always positive and highly significant. When using the shift-share strategy in col. 1 , the magnitude of the coefficient is close to that of our FE regressions. The coefficient becomes larger under the gravity-like strategy *a la* Feyrer (2009) in col. 2 even if both are not significantly different from the FE estimates. It is worth noticing that the instruments used in our IV regressions are valid. In particular, the Kleibergen-Paap F-stat of our second stage is always very large, and satisfies the Stock-Yogo critical values related to 10% maximal IV size. In addition, the F-test of the first stage is always above the critical value of 10. After instrumenting with the shift-share strategy, a 10% change in diversity induces a 5.1% change in GDP; equivalently, a one-standard-deviation change in high-skilled diversity increases GDP per capita by 2.8%, which is close to our benchmark results. As for low-skilled diversity, we find insignificant or weakly significant effects in col. 6 and 7.

We conduct additional IV regressions to deal with the endogeneity of the immigration rate and of the epidemiological term in the remaining columns of Table 2.6. As the shift-share strategy does a poor job at predicting the immigration rate,[28] we only use the gravity-like strategy *a la* Feyrer (2009). Different combinations of endogenous regressors are considered, without changing our conclusions. In all specifications, the instrumental variables are strong. Our estimates for $MD^H_{r,t}$ are robust, and the magnitude of the coefficient is similar to the FE estimates. The effect of low-skilled diversity is always insignificant from col. 8 to 10. Under some specifications, we obtain a negative and significant epidemiological effect for both college-educated and low-skilled immigrants. Again, we find no evidence of a contamination effect. On the contrary, our epidemiological results are more in line with the effect of diversity; attracting immigrants from economically and culturally distant countries is beneficial for economic growth. Overall, our IV regressions support the view that increasing birthplace diversity among college-educated immigrants causes a rise in GDP per capita at destination.

2.5 Conclusions

This paper empirically investigates the impact of multiculturalism (as measured by birthplace diversity among immigrants, birthplace polarization indices or immigration-driven epidemiological norms) on GDP per capita. To do so, we use a large sample of US states and take advantage of the availability of panel data. Compared to existing studies, our analysis relies on panel data available for a long period of fifty years, and systematically tests for skill-specific effects of cultural diversity. Using a full set of fixed effects and combining various instrumentation strategies, we find that diversity among college-educated immigrants positively affects macroeconomic performance. On the contrary, diversity among less educated immigrants has insignificant effects (neither positive nor negative), and this is not due to the higher fraction of undocumented migrants in this group. These results are highly robust to measurement, spec-

[28]The same problem arises in Alesina *et al.* (2016).

ification and instrumentation hypotheses. Furthermore, we find no evidence of a quadratic effect, or of a contamination by the bad economic conditions in poor countries.

Overall, a 10% increase in diversity among college-educated immigrants raises GDP per capita by 6.2%. Albeit non-negligible, the macroeconomic implications of diversity are limited. High-skilled diversity only explains 3.5% of the output rise between 1960 and 2010 in the US, and about 4% of the current output gap between the least and most diverse states.

Table 2.6: 2SLS regressions under different IV strategies.
(Dep= $log(y_{r,t})$)

	(1) Shift-Share $S=H$	(2) Feyrer $S=H$	(3) Feyrer $S=H$	(4) Feyrer $S=H$	(5) Feyrer $S=H$	(6) Shift-Share $S=L$	(7) Feyrer $S=L$	(8) Feyrer $S=L$	(9) Feyrer $S=L$	(10) Feyrer $S=L$
$MD_{r,t}$	0.511**	1.035***	0.853***	0.444**	0.726***	0.276*	0.186	0.089	0.105	0.096
	(0.213)	(0.308)	(0.278)	(0.200)	(0.259)	(0.157)	(0.132)	(0.142)	(0.089)	(0.151)
$m_{r,t}$	0.548*	0.878**	0.481	0.079	0.289	0.623*	0.527*	0.373	0.404	0.388
	(0.306)	(0.387)	(0.533)	(0.504)	(0.573)	(0.326)	(0.278)	(0.390)	(0.396)	(0.421)
$MY_{r,t}$			-0.122*	-0.208**	-0.215**			-0.106***	-0.098*	-0.099*
			(0.074)	(0.098)	(0.096)			(0.040)	(0.055)	(0.056)
$log(Pop_{r,t})$	-0.153**	-0.163**	-0.145*	-0.130*	-0.137*	-0.174**	-0.169**	-0.143*	-0.146*	-0.145*
	(0.074)	(0.073)	(0.076)	(0.078)	(0.078)	(0.080)	(0.076)	(0.076)	(0.083)	(0.075)
$log(Urb_{r,t})$	0.285**	0.283**	0.277**	0.276**	0.276**	0.378**	0.345**	0.307*	0.313*	0.309*
	(0.135)	(0.125)	(0.123)	(0.131)	(0.124)	(0.170)	(0.161)	(0.178)	(0.167)	(0.177)
$log(Hum_{r,t})$	0.772***	0.709***	0.705***	0.737***	0.702***	0.711***	0.776***	0.807***	0.801***	0.806***
	(0.190)	(0.199)	(0.193)	(0.186)	(0.192)	(0.210)	(0.218)	(0.211)	(0.191)	(0.210)
Endogenous regressors:										
$\widehat{MD^S_{r,t}}$	√	√	√		√	√	√	√		√
$\widehat{MY^S_{r,t}}$				√	√				√	√
$\widehat{m^S_{r,t}}$			√	√	√			√	√	√
Observations	306	306	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
K-P F-$Test^{\dagger}$	439.9	103.9	43.65	69.68	48.14	91.14	70.12	33.89	75.88	18.11
Stock Yogo	29.18/16.23	16.38/8.96	7.03/4.58	7.03/4.58	N.A	29.18/16.23	16.38/8.96	7.03/4.58	7.03/4.58	N.A

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (2.3) and includes all fixed effects. We estimate it with 2SLS and rely on two IV strategies (the augmented shift-share and the gravity-like strategy *a la* Feyrer (2009)) to instrument the birthplace diversity index. Col. 1, 2 and 3 report estimates for high-skilled diversity; col. 4, 5 and 6 report estimates for low-skilled diversity. †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). First-stage results are reported in Table A22 and A23in the Appendix. Zero-stage regressions with instrumentation *a la* Feyrer (2009) are provided in Table A20 in the Appendix. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

References

Ager, P., and Brückner, M. (2013). Cultural diversity and economic growth: Evidence form the US during the age of mass migration. *European Economic Review*, 64:76–97.

Alesina, A., Harnoss, J. and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2):101–138.

Alesina, A., and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Artuc, E., Docquier, F., Özden, C., and Parsons, C. (2015). A Global Assessment of Human Capital Mobility: The Role of Non-OECD Destinations. *World Development*, 65(1):6–26.

Ashraf, Q., and Galor, O. (2013). The Out of Africa hypothesis, Human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.

Ashraf, Q., Galor, O., and Klemp, M. (2015). The Out of Africa hypothesis of comparative development reflected by nighttime light intensity. Working Paper No. 2014-4, Brown University, Departments of Economics.

Beine, M., Docquier, F., and Schiff, M. (2013). International migration, transfer of norms and home country fertility. *Canadian Journal of Economics*, 46(4):1406–1430.

Bertoli, S. and Marchetta, F. (2015). Bringing it all back home–return migration and fertility choices. *World Development*, 65:27–40.

Boeheim, R., Horvath, G., and Mayr, K. (2012). Birthplace diversity of the workforce and productivity spill-overs in firms. WIFO Working PapersNo. 438.

Borjas, G.J. (2015). Immigration and Globalization: A review Essay. *Journal of Economic Literature*, 53(4):961–974.

Borjas, G.J. (2016). The labor supply of undocumented immigrants. NBER working paper No. 22102. National Bureau of Economic Research Cambridge, MA.

Collier, P. (2013). Exodus: How migration is changing our World. Oxford University Press (309 p.).

Dorn, D. (2009). Essays on Inequality, Spatial Interaction, and the Demand for Skills. Dissertation University of St. Gallen no. 3613, September 2009.

Feyrer, J. (2009). Trade and income: Exploiting time series in geography. NBER working paper No. 14910. National Bureau of Economic Research Cambridge, MA.

Fulford, S., Petkov, I., and Schiantarelli, F. (2015). Does it matter where you came from? Ancestry composition and economic performance of U.S. counties, 1850-2010. Boston College Working Papers in Economics No. 875.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: Assessing the impacts of cultural and language diversity in the national hockey league. *The Review of Economics and Statistics*, 95(1):302–314.

Lodigiani, E. and Salomone, S. (2012). Migration-induced transfers of norms: The case of female political empowerment. *IRES Discussion Papers 2012–1.*

McKeown, A. (2004). Global migration 1846-1940. *Journal of World History*, 15(2):155–189.

Montalvo, J.G., and Reynal-Querol, M. (2003). Religious polarization and economic development *Economics Letters*, 80(2):201–210.

Montalvo, J.G., and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.

Nickel, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Ottaviano, G., and Peri, G. (2006). The economic value of cultural diversity: Evidence from U.S. cities. *Journal of Economic Geography*, 6(1):9–44.

Özden, Ç., Parsons C. R., Schiff, M., and Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960-2000. *World Bank Economic Review*, 25(1):12–56.

Ozgen, C., Peters, C., Niebuhr, A., Nijkamp, P. and Poot, J. (2014). Does cultural diversity of migrants employees affect innovation? *International Migration Review*, 48(9):377–416.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2014). Does labor diversity affect firm productivity? *European Economic Review*, 66:144–179.

Putterman, L., and Weill, D.N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *The Quarterly Journal of Economics*, 125(4):1627–1682.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek M. (2015). Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Santos Silva, J. M. C., and Tenreyro S. (2006). "The Log of Gravity". *Review of Economics and Statistics*, 88(4):641–658.

Spilimbergo, A. (2009). Democracy and foreign education. *The American Economic Review*, 99(1):528–543.

Spolaore E, and Wacziarg R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2):469–592.

Spolaore E, and Wacziarg R. (2015). Ancestry, Language and Culture. NBER Working Paper No. 21242. National Bureau of Economic Research Cambridge, MA.

Suedekum, J., Wolf, K., and Blien, U. (2014). Cultural diversity and local labour markets. *Regional Studies*, 48(1):173–191.

Appendix to chapter 2

Table A7: Variables: Source and definition.

Variable	Description	Definition	Source
$y_{r,t}$	Gross Domestic product	Gross domestic product (GDP) per capita.	Bureau of Economic Analysis.
$TD^S_{r,t}$	Birthplace diversity among residents	Probability that two randomly-drawn residents in region r have different countries of birth.	Authors' calculation on IPUMS-US data.
$MD^S_{r,t}$	Birthplace diversity among immigrants	Probability that two randomly-drawn immigrants in region r have different countries of birth.	Authors' calculation on IPUMPS-US data.
$MY^S_{r,t}$	Immigration-driven norm among immigrants	Weighted average outcome y in immigrant's origin countries where the weights are immigrants' share in the total immigrants' population in the destination region r.	Authors' calculation on IPUMS-US data.
$TP^S_{r,t}$	Polarization index among residents	Index that captures how the birthplace distribution in a population is far from the bimodal distribution.	Authors' calculation on IPUMS-US data.
$MP^S_{r,t}$	Polarization index among immigrants	Index that captures how the birthplace distribution in a immigrant's population is far from the bimodal distribution.	Authors' calculation on IPUMS-US data.
$k^S_{i,r,t}$	Share of immigrants	Number of individuals born in country i and living in region r as percentage of the total population of region r at year t.	IPUMS-US data.
$\widehat{k}^S_{i,r,t}$	Share of immigrants	Share of immigrants from origin country i in the total immigrant population of region r.	IPUMS-US data.
$m^S_{r,t}$	Immigration rate	Ratio of the total stock of foreign-born individuals to the total population of region r at year t.	Author's calculation on IPUMS-US.
$Hum_{r,t}$	Average education	Average education level.	Authors' calculation IPUMS data.
$Pop_{r,t}$	Population	Population of region r at year t.	IPUMS-US data.
$Urb_{r,t}$	Urbanization	Urban Percentage of the Population for States.	U.S. Census Bureau.
Gravity model:			
$Stock_{i,r,t}$	Stock of immigrants	Number of individuals born in country i and living in region r at year t.	IPUMS-US data.
$Distance_{i,r}$	Distance	Great-circle distance between the capital city of the origin countries i and the capital of the destination region r for US states.	Authors' calculation.
$Bord_{Canada,r}$	Common border	Dummy equal to 1 if Canada and State r share a common border and 0 otherwise.	Authors' elaboration.
$Bord_{Mexico,r}$	Common border	Dummy equal to 1 if Mexico and State r share a common border and 0 otherwise.	Authors' calculation.

Source: Authors' calculation.

Table A8: List of origin countries (195).

Afghanistan, Albania, Algeria, Andorra, Angola, Antigua-Barbuda, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bhutan, Bolivia, Bosnia, Botswana, Brazil, Brunei, Bulgaria, Burkina Faso, Burma (Myanmar), Burundi, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Chad, Chile, China, China, Hong Kong SAR, China, Macao SAR, Colombia, Comoros, Congo, Dem. Rep. of the, Congo, Rep. of the, Costa Rica, Cote d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Djibouti, Dominica, Dominican Republic, East Timor, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, Gabon, Gambia, Georgia, Germany, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana/British Guiana, Haiti, Holy See (Vatican City), Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, Korea, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Liechtenstein, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mexico, Micronesia, Moldova, Monaco, Mongolia, Morocco, Mozambique, Namibia, Nauru, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Occupied Palestinian Territory, Oman, Pakistan, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Romania, Russia, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia and Montenegro, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Tuvalu, Uganda, Ukraine, United Arab Emirates, United Kingdom, Unites States, Uruguay, Uzbekistan, Vanuatu, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

Table A9: List of US States (51) and descriptives statistics

States	$log(y_{r,t})$	$MD^A_{r,t}$	$MD^H_{r,t}$	$MD^L_{r,t}$	$m^A_{r,t}$	$log(Pop_{r,t})$	$log(Urb_{r,t})$	$log(Hum_{r,t})$
Alabama	9,256	0,900	0,923	0,846	0,021	14,715	4,060	1,722
Alaska	10,053	0,894	0,880	0,882	0,064	12,561	4,071	1,855
Arizona	9,484	0,690	0,906	0,554	0,107	14,440	4,428	1,806
Arkansas	9,209	0,871	0,914	0,811	0,024	14,154	3,930	1,706
California	9,731	0,845	0,941	0,734	0,221	16,618	4,519	1,839
Colorado	9,616	0,872	0,934	0,778	0,071	14,490	4,393	1,888
Connecticut	9,759	0,940	0,952	0,928	0,116	14,518	4,402	1,856
Delaware	9,847	0,930	0,932	0,889	0,055	12,919	4,303	1,827
District of Columbia	10,564	0,960	0,968	0,903	0,106	13,016	4,605	1,893
Florida	9,424	0,869	0,889	0,856	0,143	15,675	4,431	1,794
Georgia	9,458	0,909	0,939	0,854	0,052	15,174	4,164	1,741
Hawaii	9,657	0,745	0,828	0,711	0,172	13,347	4,457	1,846
Idaho	9,336	0,811	0,869	0,714	0,046	13,299	4,057	1,824
Illinois	9,675	0,883	0,947	0,792	0,109	15,807	4,438	1,816
Indiana	9,501	0,917	0,950	0,852	0,032	15,066	4,197	1,784
Iowa	9,476	0,909	0,945	0,856	0,029	14,379	4,077	1,826
Kansas	9,470	0,878	0,940	0,791	0,044	14,237	4,219	1,849
Kentucky	9,369	0,905	0,930	0,859	0,019	14,640	3,953	1,690
Louisiana	9,501	0,948	0,958	0,925	0,028	14,728	4,229	1,721
Maine	9,319	0,656	0,796	0,590	0,047	13,481	3,812	1,802
Maryland	9,592	0,955	0,960	0,938	0,088	14,872	4,389	1,836
Massachusetts	9,700	0,930	0,951	0,911	0,122	15,148	4,461	1,873
Michigan	9,561	0,927	0,928	0,910	0,060	15,568	4,290	1,809
Minnesota	9,580	0,938	0,950	0,907	0,046	14,789	4,222	1,855
Mississippi	9,119	0,917	0,898	0,879	0,015	14,242	3,820	1,701
Missouri	9,490	0,939	0,950	0,913	0,028	14,969	4,232	1,785
Montana	9,344	0,873	0,877	0,859	0,029	13,113	3,973	1,845
Nebraska	9,510	0,895	0,937	0,838	0,039	13,792	4,164	1,846
Nevada	9,759	0,863	0,923	0,802	0,125	13,342	4,439	1,818
New Hampshire	9,470	0,809	0,887	0,772	0,057	13,315	4,028	1,843
New Jersey	9,717	0,949	0,951	0,941	0,159	15,394	4,509	1,832
New Mexico	9,477	0,660	0,893	0,509	0,071	13,655	4,278	1,796
New York	9,805	0,954	0,964	0,944	0,189	16,276	4,453	1,827
North Carolina	9,475	0,905	0,948	0,845	0,039	15,232	3,929	1,737
North Dakota	9,362	0,873	0,834	0,857	0,029	12,886	3,889	1,817
Ohio	9,531	0,951	0,951	0,940	0,037	15,732	4,320	1,799
Oklahoma	9,360	0,883	0,942	0,796	0,036	14,445	4,193	1,798
Oregon	9,511	0,866	0,920	0,788	0,072	14,342	4,262	1,852
Pennsylvania	9,505	0,945	0,952	0,933	0,046	15,844	4,287	1,796
Rhode Island	9,478	0,888	0,931	0,858	0,109	13,346	4,477	1,790
South Carolina	9,273	0,903	0,924	0,857	0,028	14,548	3,981	1,716
South Dakota	9,326	0,911	0,901	0,893	0,021	12,971	3,868	1,809
Tennessee	9,375	0,916	0,943	0,863	0,025	14,920	4,099	1,720
Texas	9,569	0,610	0,917	0,452	0,113	16,091	4,385	1,764
Utah	9,450	0,883	0,925	0,822	0,066	13,739	4,432	1,880
Vermont	9,387	0,754	0,852	0,679	0,053	12,690	3,570	1,839
Virginia	9,560	0,953	0,952	0,940	0,070	15,132	4,201	1,802
Washington	9,709	0,897	0,912	0,853	0,098	14,862	4,331	1,870
West Virginia	9,248	0,933	0,915	0,910	0,013	13,958	3,700	1,700
Wisconsin	9,504	0,909	0,946	0,849	0,038	14,910	4,194	1,818
Wyoming	9,755	0,885	0,894	0,818	0,033	12,513	4,134	1,844

Note: Average from 1960 to 2010. Source: Authors' elaboration on IPUMS-US data.

Table A10: Robustness of FE regressions for high-skilled diversity. Alternative sub-samples (Dep= $log(y_{r,t})$)

	(1) Full Sample	(2) 1970-2000	(3) No Top5	(4) No Bot5	(5) No Mex	(6) No $Q1$
$MD^H_{r,t}$	0.616***	0.870***	0.725***	0.672***	0.630***	0.596**
	(0.160)	(0.321)	(0.174)	(0.170)	(0.170)	(0.288)
$m^H_{r,t}$	0.614*	1.140**	1.317**	0.613*	0.541	0.765**
	(0.315)	(0.459)	(0.529)	(0.323)	(0.397)	(0.365)
$log(Pop_{r,t})$	-0.155**	0.002	-0.187**	-0.160**	-0.158*	-0.182**
	(0.075)	(0.075)	(0.082)	(0.073)	(0.088)	(0.074)
$log(Urb_{r,t})$	0.285**	0.290	0.260*	0.300**	0.295**	0.198
	(0.135)	(0.187)	(0.140)	(0.147)	(0.138)	(0.151)
$log(Hum_{r,t})$	0.759***	1.251***	0.692***	0.945***	0.731***	0.797***
	(0.197)	(0.310)	(0.213)	(0.183)	(0.224)	(0.233)
Constant	7.348***	4.309***	7.870***	7.030***	7.373***	8.107***
	(1.262)	(1.584)	(1.387)	(1.250)	(1.398)	(1.313)
Observations	306	204	276	276	282	228
Nb. states	51	51	46	46	47	38
R-squared	0.993	0.990	0.993	0.993	0.993	0.995
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq.(2.3) and includes all fixed effects. Col. 1 reports the results from Table 2.2. In col. 2, we exclude observations for the years 1960 and 2010. In col. 3 and 4, we exclude the five US states with the greatest or smallest immigration shares. In col. 5, we exclude US states located on the US-Mexican border. In col. 6, we exclude the lowest quartile in terms of immigrant population. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

Table A11: Robustness of FE regressions for low-skilled diversity. Alternative sub-samples (Dep= $log(y_{r,t})$)

	(1) Full Sample	(2) 1970-2000	(3) No Top5	(4) No Bot5	(5) No Mex	(6) No $Q1$
$MD^L_{r,t}$	0.141	0.130	0.228**	0.128	0.109	0.015
	(0.086)	(0.097)	(0.093)	(0.092)	(0.091)	(0.115)
$m^L_{r,t}$	0.481*	0.691**	0.938**	0.448	0.474	0.647**
	(0.282)	(0.272)	(0.418)	(0.288)	(0.392)	(0.276)
$log(Pop_{r,t})$	-0.166**	0.004	-0.200**	-0.169**	-0.157	-0.197**
	(0.081)	(0.086)	(0.090)	(0.080)	(0.095)	(0.076)
$log(Urb_{r,t})$	0.329**	0.323	0.315*	0.341*	0.313*	0.360**
	(0.163)	(0.200)	(0.165)	(0.182)	(0.168)	(0.166)
$log(Hum_{r,t})$	0.807***	1.309***	0.743***	0.964***	0.785***	0.919***
	(0.205)	(0.343)	(0.212)	(0.204)	(0.215)	(0.264)
Constant	7.662***	4.738***	8.183***	7.439***	7.660***	7.971***
	(1.263)	(1.514)	(1.370)	(1.259)	(1.429)	(1.158)
Observations	306	204	276	276	282	228
Nb. states	51	51	46	46	47	38
R-squared	0.993	0.989	0.993	0.993	0.993	0.996
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq.(2.3) and includes all fixed effects. Col. 1 reports the results from Table 2.2. In col. 2, we exclude observations for the years 1960 and 2010. In col. 3 and 4, we exclude the five US states with the greatest or smallest immigration shares. In col. 5, we exclude US states located on the US-Mexican border. In col. 6, we exclude the lowest quartile in terms of immigration rate. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

Table A12: Robustness of FE regressions for high-skilled diversity.
Ten largest US immigrants group in 2010 (Dep= $log(y_{r,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
i	Mexico	India	Philippines	China	Vietnam	El Salvador	Cuba	Korea	Dominican Rep.	Guatemala	10 largest
$MD^H_{r,t}$	0.649***	0.616***	0.602***	0.577***	0.615***	0.613***	0.627***	0.617***	0.607***	0.621***	0.617***
	(0.176)	(0.161)	(0.164)	(0.149)	(0.160)	(0.156)	(0.158)	(0.160)	(0.156)	(0.163)	(0.169)
$m^H_{r,t}$	0.673**	0.743**	0.626*	0.638**	0.615*	0.500	0.631**	0.628*	0.550	0.649**	0.726*
	(0.316)	(0.349)	(0.315)	(0.311)	(0.315)	(0.303)	(0.301)	(0.336)	(0.338)	(0.322)	(0.365)
$\widehat{k}^H_{i,r,t}$	0.214	-0.628*	-0.126	0.403	-0.035	2.033	0.063	-0.158	0.354	-0.813	
	(0.260)	(0.352)	(0.199)	(0.280)	(0.458)	(1.483)	(0.237)	(0.425)	(0.712)	(1.313)	
$log(Pop_{r,t})$	-0.166**	-0.174**	-0.150**	-0.144*	-0.155**	-0.157**	-0.155**	-0.156**	-0.151*	-0.156**	-0.169**
	(0.078)	(0.075)	(0.074)	(0.079)	(0.076)	(0.074)	(0.075)	(0.076)	(0.079)	(0.075)	(0.083)
$log(Urb_{r,t})$	0.282**	0.278**	0.283**	0.292**	0.286**	0.271*	0.285**	0.289**	0.296**	0.287**	0.282*
	(0.136)	(0.139)	(0.139)	(0.134)	(0.138)	(0.139)	(0.136)	(0.136)	(0.145)	(0.138)	(0.162)
$log(Hum_{r,t})$	0.807***	0.852***	0.743***	0.766***	0.758***	0.760***	0.766***	0.753***	0.752***	0.759***	0.895***
	(0.218)	(0.196)	(0.198)	(0.198)	(0.205)	(0.196)	(0.202)	(0.199)	(0.198)	(0.195)	(0.233)
Constant	7.404***	7.506***	7.328***	7.173***	7.349***	7.442***	7.327***	7.352***	7.263***	7.356***	7.341***
	(1.263)	(1.212)	(1.248)	(1.338)	(1.266)	(1.240)	(1.272)	(1.266)	(1.354)	(1.262)	(1.360)
Observations	306	306	306	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. $\widehat{k}^H_{i,r,t}$ is the share of high-skilled immigrants from origin country i in the total immigrant population of state r. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. Col. 11 includes $\widehat{k}^H_{i,r,t}$ for the 10 largest immigrant groups in the US. Coefficients for all the countries are not reported in col. 11 for space limitations.

Table A13: Robustness of FE regressions for low-skilled diversity.
Ten largest US immigrants group in 2010 (Dep= $log(y_{r,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
i	Mexico	India	Philippines	China	Vietnam	El Salvador	Cuba	Korea	Dominican Rep.	Guatemala	10 largest
$MD^L_{r,t}$	0.294**	0.149*	0.159	0.213**	0.127	0.133	0.131	0.148	0.147*	0.142*	0.276**
	(0.111)	(0.086)	(0.097)	(0.093)	(0.093)	(0.082)	(0.089)	(0.090)	(0.086)	(0.085)	(0.104)
$m^L_{r,t}$	0.652**	0.454*	0.252	0.483*	0.484*	0.211	0.446	0.482*	0.500*	0.443	0.058
	(0.290)	(0.266)	(0.266)	(0.260)	(0.282)	(0.249)	(0.283)	(0.281)	(0.293)	(0.291)	(0.274)
$\widehat{k}^L_{i,r,t}$	0.173*	-1.781	-1.472*	-1.394*	0.250	0.620***	-0.232	-0.103	-0.087	0.351	
	(0.089)	(1.159)	(0.803)	(0.831)	(0.364)	(0.209)	(0.209)	(0.295)	(0.414)	(0.399)	
$log(Pop_{r,t})$	-0.187**	-0.172**	-0.151**	-0.190**	-0.161*	-0.127	-0.164**	-0.166**	-0.169*	-0.162*	-0.129*
	(0.078)	(0.079)	(0.071)	(0.075)	(0.084)	(0.077)	(0.080)	(0.081)	(0.086)	(0.082)	(0.069)
$log(Urb_{r,t})$	0.335*	0.307*	0.308**	0.317**	0.312*	0.320*	0.325**	0.336**	0.328*	0.329**	0.256*
	(0.167)	(0.163)	(0.147)	(0.153)	(0.174)	(0.162)	(0.161)	(0.153)	(0.164)	(0.162)	(0.149)
$log(Hum_{r,t})$	0.792***	0.839***	0.681***	0.810***	0.804***	0.733***	0.808***	0.801***	0.808***	0.777***	0.622***
	(0.203)	(0.209)	(0.217)	(0.207)	(0.202)	(0.211)	(0.200)	(0.204)	(0.206)	(0.207)	(0.218)
Constant	7.804***	7.786***	7.768***	7.999***	7.683***	7.292***	7.655***	7.637***	7.698***	7.659***	7.665***
	(1.273)	(1.229)	(1.166)	(1.172)	(1.267)	(1.097)	(1.254)	(1.262)	(1.326)	(1.261)	(0.975)
Observations	306	306	306	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.994
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. $\widehat{k}^L_{i,r,t}$ is the share of low-skilled immigrants from origin country i in the total immigrant population of state r. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. Col. 11 includes $\widehat{k}^L_{i,r,t}$ for the 10 largest immigrant groups in the US. Coefficients for all the countries are not reported in col. 11 for space limitations.

Table A14: Robustness of FE estimates to alternative specifications. Results by skill group (Dep= $log(y_{r,t})$)

	(1) Quadratic. $S=H$	(2) Polarization $S=H$	(3) Quadratic $S=L$	(4) Polarization $S=L$
$MD^S_{r,t}$	-0.131		0.705**	
	(1.954)		(0.293)	
$(MD^S_{r,t})^2$	0.453		-0.391*	
	(1.202)		(0.218)	
$MP^S_{r,t}$		-0.291***		-0.025
		(0.090)		(0.072)
$m^S_{r,t}$	0.622*	0.596*	0.504*	0.352
	(0.314)	(0.306)	(0.281)	(0.271)
$log(Pop_{r,t})$	-0.154**	-0.154**	-0.164**	-0.159*
	(0.075)	(0.074)	(0.081)	(0.082)
$log(Urb_{r,t})$	0.279*	0.254*	0.347**	0.282
	(0.143)	(0.142)	(0.159)	(0.170)
$log(Hum_{r,t})$	0.758***	0.748***	0.799***	0.894***
	(0.197)	(0.196)	(0.206)	(0.202)
Constant	7.666***	8.121***	7.388***	7.762***
	(1.571)	(1.224)	(1.284)	(1.271)
Observations	306	306	306	306
Nb. states	51	51	51	51
R-squared	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (2.3) and includes all fixed effects. Col. 1 and 3 tests for a quadratic specification in birthplace diversity with $MD^H_{r,t}$ and $(MD^H_{r,t})^2$. In col. 2 and 4, we replace birthplace diversity by a polarization index ($MP^H_{r,t}$). The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

Table A15: Robustness of Pooled OLS, FE and IV regressions without controls.
Results by skill group (Dep= $log(y_{r,t})$)

	(1) OLS $S=H$	(2) Fixed-effects $S=H$	(3) Shift-Share $S=H$	(4) Feyrer $S=H$	(5) OLS $S=L$	(6) Fixed-effects $S=L$	(7) Shift-Share $S=L$	(8) Feyrer $S=L$
$MD^S_{r,t}$	7.595***	0.664***	0.418*	1.079***	0.343	0.147**	0.009	0.130
	(0.909)	(0.209)	(0.234)	(0.376)	(0.498)	(0.063)	(0.120)	(0.192)
$m^S_{r,t}$	9.912***	-0.019	-0.168	0.231	6.963***	-0.123	-0.290	-0.143
	(1.421)	(0.452)	(0.452)	(0.529)	(1.053)	(0.314)	(0.334)	(0.468)
Observations	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51
R-squared	0.452	0.992	0.992	0.991	0.236	0.991	0.991	0.991
Time fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
States fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
K-P F-$Test^\dagger$			383.45	107.45			100.82	47.31
Stock Yogo			16.38/ 8.96	16.38/ 8.96			16.38/ 8.96	16.38/ 8.96

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (2.3). The sample includes the 50 US states and the District of Columbia from 1960 to 2010. We estimate 2SLS relying on two IV strategies (the augmented shift-share and the gravity-like strategy *a la* Feyrer (2009)) to instrument the birthplace diversity index. †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size).

Table A16: Robustness of FE regressions to alternative educational levels (Dep= $log(y_{r,t})$)

	(1) $S=H$	(2) $S=Ph.D$	(3) $S=Tertiary$	(4) $S=L$	(5) $S=No\ school$	(6) $S=Primary$	(7) $S=Secondary$
$MD^S_{r,t}$	0.616***	0.262**	0.369***	0.141	0.038	0.033	0.120
	(0.160)	(0.103)	(0.136)	(0.086)	(0.032)	(0.061)	(0.101)
$m^S_{r,t}$	0.614*	0.256	0.372	0.481*	0.152	0.165	0.432
	(0.315)	(0.266)	(0.298)	(0.282)	(0.100)	(0.101)	(0.301)
$log(Pop_{r,t})$	-0.155**	-0.158**	-0.141*	-0.166**	-0.167**	-0.160**	-0.149*
	(0.075)	(0.077)	(0.077)	(0.081)	(0.076)	(0.076)	(0.080)
$log(Urb_{r,t})$	0.285**	0.287*	0.253*	0.329**	0.207	0.287	0.311*
	(0.135)	(0.145)	(0.148)	(0.163)	(0.168)	(0.173)	(0.163)
$log(Hum_{r,t})$	0.759***	0.763***	0.779***	0.807***	0.843***	1.004***	0.831***
	(0.197)	(0.205)	(0.206)	(0.205)	(0.210)	(0.206)	(0.209)
Constant	7.348***	7.702***	7.489***	7.662***	8.205***	7.552***	7.497***
	(1.262)	(1.244)	(1.290)	(1.263)	(1.342)	(1.197)	(1.300)
Observations	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Col 1 and 4 report our benchmark specifications from Table 2.

Table A17: Robustness of FE regressions.
Results by legal status and skill group (Dep= $log(y_{r,t})$)

	(1) $S=H$ All	(2) $S=H$ Legal	(3) $S=H$ Undoc.	(4) $S=L$ All	(5) $S=L$ Legal	(6) $S=L$ Undoc.
$MD^S_{r,t}$	0.813*	1.009**	-0.153	0.038	0.043	-0.017
	(0.438)	(0.473)	(0.127)	(0.100)	(0.112)	(0.045)
$m^S_{r,t}$	0.842*	0.959*	4.426**	0.957**	1.107**	3.482*
	(0.431)	(0.535)	(2.140)	(0.377)	(0.424)	(1.795)
$ln(Population_{s,t})$	-0.029	-0.028	0.013	-0.074	-0.057	-0.088
	(0.092)	(0.094)	(0.088)	(0.105)	(0.099)	(0.112)
$ln(Urban_{s,t})$	0.129	0.129	0.050	0.093	0.090	0.079
	(0.164)	(0.166)	(0.172)	(0.167)	(0.166)	(0.180)
$ln(College_{s,t})$	2.043***	1.997***	2.296***	2.508***	2.418***	2.671***
	(0.633)	(0.622)	(0.611)	(0.690)	(0.676)	(0.724)
Constant	4.762***	4.650**	4.839***	5.450***	5.366***	5.497***
	(1.723)	(1.758)	(1.647)	(1.630)	(1.634)	(1.706)
Observations	204	204	204	204	204	204
Nb. states	51	51	51	51	51	51
R-squared	0.979	0.979	0.979	0.979	0.979	0.979
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the state level. Col. 1 and 4 report the coefficient of the benchmark sample over the 1980-2010 period and for high-skilled and low-skilled immigrants, respectively. In col. 2 and 5, the diversity indices are computed for the legal immigrant population only. In col. 3 and 6, we use the undocumented immigrant population only.

Table A18: Robustness of FE and IV regressions to spatial scale.
Results by skill group at the Commuting Zones level (Dep= $log(Wage_{CZs,t})$)

	(1) Fixed-effects $S = H$	(2) Fixed-effects $S = L$	(3) Shift-Share $S = H$	(4) Shift-Share $S = L$
$MD^S_{r,t}$	0.319**	0.171**	0.372***	-0.317
	(0.150)	(0.076)	(0.128)	(0.207)
$m^S_{r,t}$	1.907***	0.997***	1.913***	0.439
	(0.567)	(0.350)	(0.561)	(0.438)
Constant	5.395***	5.528***		
	(0.110)	(0.063)		
Observations	3,688	3,688	3,688	3,688
R-squared	0.895	0.894	0.895	0.891
Number of CZs	741	741	741	741
Time fixed effects	Yes	Yes	Yes	Yes
CZs fixed effects	Yes	Yes	Yes	Yes
F-test (IV)			2305	501.3

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the Commuting zones level (CZs). The specification includes all fixed effects. The sample includes the 50 US states and the District of Columbia from 1970 to 2010. The dependent variable $= log(Wage_{CZs,t})$ is logarithm of the average wage of white US natives between 40-50 which is not affected by discrimination, following Ottaviano and Peri (2006). Commuting zones are computed following Dorn (2009).

Table A19: Pearson correlations between diversity measures

Variables	$TD^A_{r,t}$	$MD^A_{r,t}$	$TP^A_{r,t}$	$MP^A_{r,t}$	$MD^{A,G}_{r,t}$	$MY^A_{r,t}$	$m^A_{r,t}$
$TD^A_{r,t}$	1.000						
$MD^A_{r,t}$	-0.194***	1.000					
$TP^A_{r,t}$	0.987***	-0.244***	1.000				
$MP^A_{r,t}$	0.146***	-0.892***	0.190***	1.000			
$MD^{A,G}_{r,t}$	0.314***	0.393***	0.298***	-0.449***	1.000		
$MY^A_{r,t}$	-0.261***	-0.166***	-0.253***	0.189***	-0.182***	1.000	
$m^A_{r,t}$	0.998***	-0.196***	0.977***	0.152***	0.306***	-0.256**	1.000
Variables	$TD^H_{r,t}$	$MD^H_{r,t}$	$TP^H_{r,t}$	$MP^H_{r,t}$	$MD^{H,G}_{r,t}$	$MY^H_{r,t}$	$m^H_{r,t}$
$TD^H_{r,t}$	1.000						
$MD^H_{r,t}$	0.169***	1.000					
$TP^H_{r,t}$	0.990***	0.178***	1.000				
$MP^H_{r,t}$	-0.237***	-0.968***	-0.253***	1.000			
$MD^{H,G}_{r,t}$	0.430***	0.685***	0.439***	-0.733***	1.000		
$MY^H_{r,t}$	-0.128**	-0.189***	-0.116**	0.186***	-0.094	1.000	
$m^H_{r,t}$	0.999***	0.158***	0.984***	-0.224***	0.422***	-0.130**	1.000
Variables	$TD^L_{r,t}$	$MD^L_{r,t}$	$TP^L_{r,t}$	$MP^L_{r,t}$	$MD^{L,G}_{r,t}$	$MY^L_{r,t}$	$m^L_{r,t}$
$TD^L_{r,t}$	1.000						
$MD^L_{r,t}$	-0.340***	1.000					
$TP^L_{r,t}$	0.986***	-0.413***	1.000				
$MP^L_{r,t}$	0.238***	-0.828***	0.297***	1.000			
$MD^{L,G}_{r,t}$	0.194***	0.377***	0.156***	-0.364***	1.000		
$MY^L_{r,t}$	-0.322***	-0.148**	-0.313***	0.148**	-0.283***	1.000	
$m^L_{r,t}$	0.995***	-0.349***	0.972***	0.246***	0.178***	-0.303***	1.000

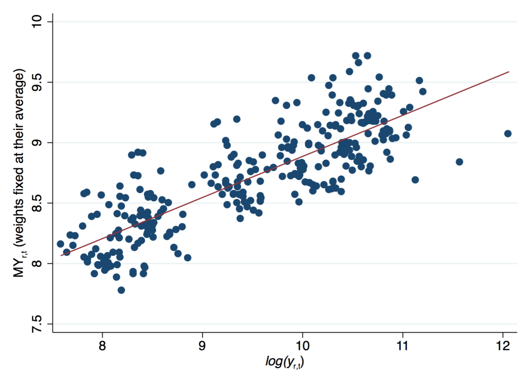
Notes: *** p<0.01, ** p<0.05. Source: Authors' elaboration on IPUMS data.

Table A20: Zero-stage estimates (PPML): gravity model *a la* Feyrer (2009)

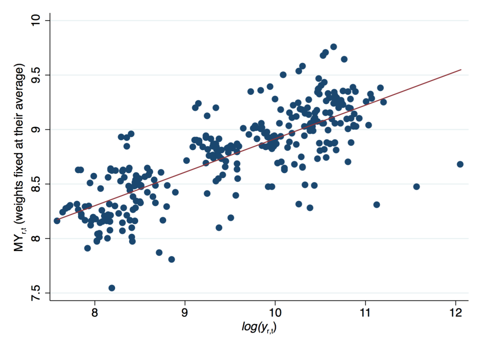
	(1)	(2)	(3)
	$S = A$	$S = H$	$S = L$
	$log(Stock_{i,r,t})$	$log(Stock_{i,r,t})$	$log(Stock_{i,r,t})$
$log(Dist_{i,r}) \times I_{1960}$	-1.666***	-1.480***	-1.741***
	(0.32)	(0.31)	(0.34)
$log(Dist_{i,r}) \times I_{1970}$	-1.786***	-1.463***	-1.954***
	(0.33)	(0.33)	(0.35)
$log(Dist_{i,r}) \times I_{1980}$	-1.760***	-1.349***	-2.033***
	(0.35)	(0.35)	(0.36)
$log(Dist_{i,r}) \times I_{1990}$	-1.733***	-1.280***	-2.112***
	(0.35)	(0.35)	(0.36)
$log(Dist_{i,r}) \times I_{2000}$	-1.821***	-1.241***	-2.250***
	(0.35)	(0.35)	(0.36)
$log(Dist_{i,r}) \times I_{2010}$	-1.827***	-1.276***	-2.309***
	(0.35)	(0.35)	(0.36)
$Bord_{Canada,r}$	3.605***	2.710***	4.352***
	(0.84)	(0.70)	(0.95)
$Bord_{Mexico,r}$	1.153***	0.999***	1.240***
	(0.20)	(0.24)	(0.21)
Constant	18.808***	15.445***	19.162***
	(2.96)	(2.89)	(3.13)
Observations	59364	59364	59364
R-squared	0.884	0.777	0.907
Year dummies	Yes	Yes	Yes
Origin dummies	Yes	Yes	Yes
Destination dummies	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Col.1 shows the results of the stocks of all migrants. Columns 2 and 3 illustrate the results for the college and low educated immigrants respectively. Standard errors in parentheses adjusted for clustering at the state/country-pair level. Source: Authors' elaboration on IPUMS data.

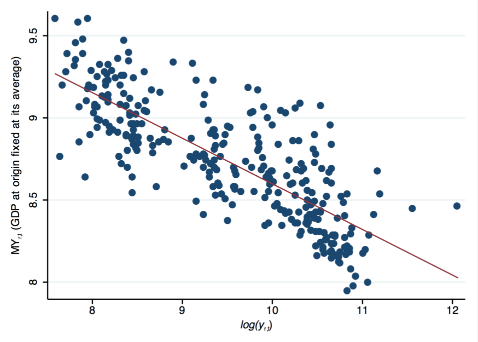
Figure A7: Cross-state correlation between the epidemiological term and GDP per capita (in logs)



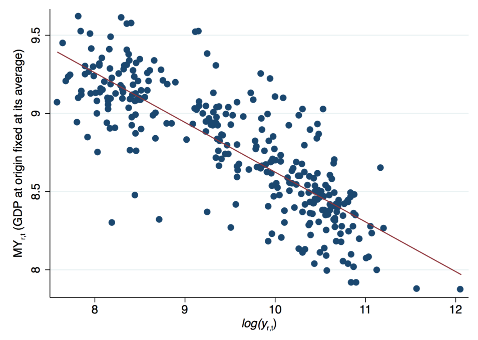
High-skilled



Low-skilled



High-skilled



Low-skilled

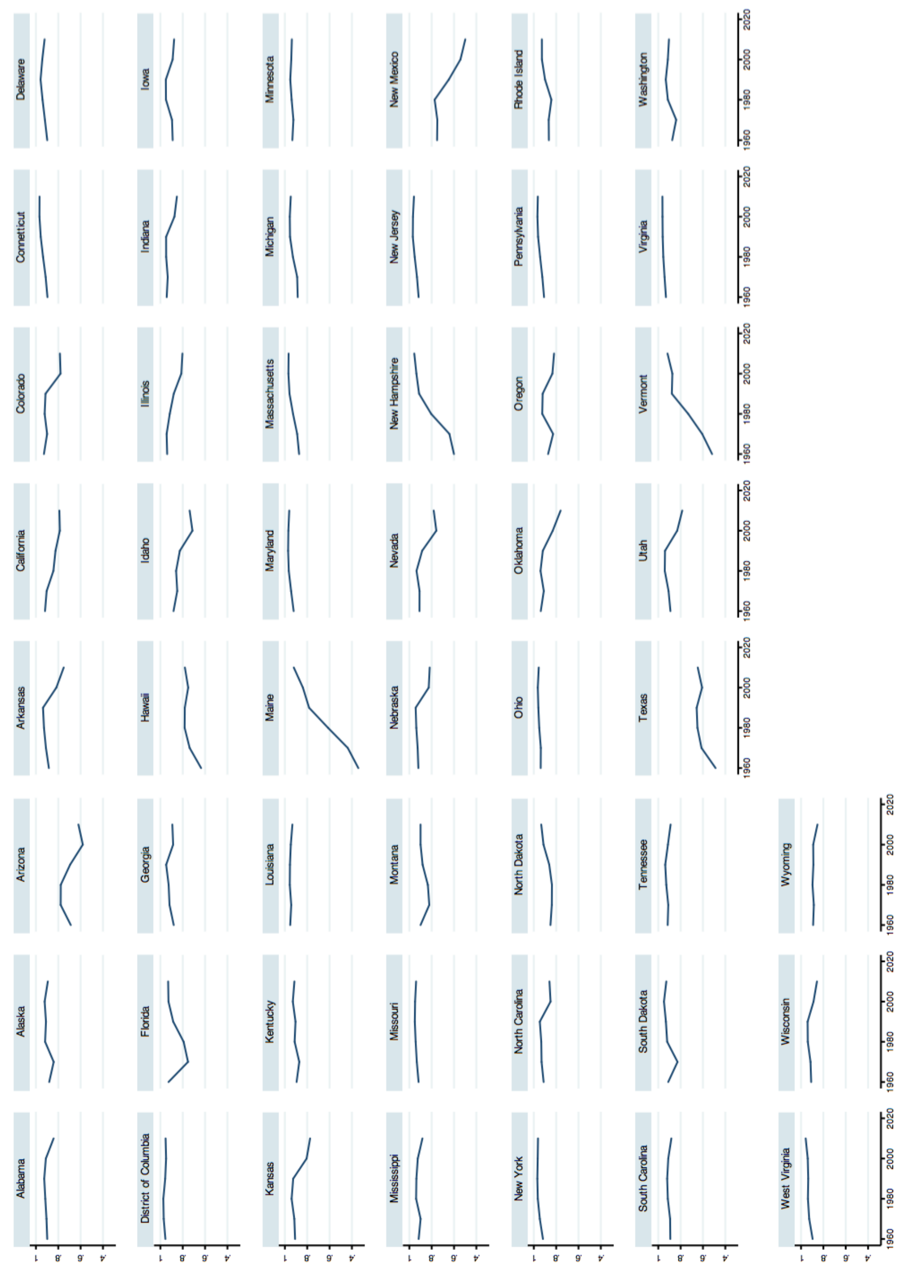
Source: Authors' elaboration on IPUMS data.

Table A21: Robustness of FE estimates to alternative definitions of the epidemiological term.
Results by skill group (Dep= $log(y_{r,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$S = H$	$S = H$	$S = H$	$S = H$	$S = L$	$S = L$	$S = L$	$S = L$
$MD^S_{r,t}$	0.531***	0.618***	0.538***	0.725***	0.104	0.141	0.101	0.006
	(0.159)	(0.162)	(0.157)	(0.249)	(0.085)	(0.085)	(0.086)	(0.085)
$MY^S_{r,t}$	-0.133*				-0.104**			
	(0.069)				(0.042)			
$MY^S_{r,t}$ (Const. $\widehat{k}^S_{i,r,t}$)		-0.120				0.001		
		(0.255)				(0.234)		
$MY^S_{r,t}$ (Const. $y_{i,t}$)			-0.110				-0.092**	
			(0.075)				(0.044)	
$MY^S_{r,t}$ ($y_{i,Entry}$)				-0.111***				-0.146**
				(0.039)				(0.057)
$m^S_{r,t}$	0.388	0.582*	0.441	0.539	0.412	0.481*	0.443	0.530**
	(0.366)	(0.300)	(0.368)	(0.328)	(0.283)	(0.273)	(0.285)	(0.241)
$log(Pop_{r,t})$	-0.144*	-0.157**	-0.143*	-0.080	-0.146*	-0.166**	-0.151*	-0.071
	(0.080)	(0.077)	(0.082)	(0.065)	(0.082)	(0.078)	(0.084)	(0.075)
$log(Urb_{r,t})$	0.282**	0.295**	0.273**	0.156	0.312*	0.329**	0.317*	0.194
	(0.135)	(0.130)	(0.134)	(0.138)	(0.173)	(0.163)	(0.172)	(0.169)
$log(Hum_{r,t})$	0.744***	0.740***	0.762***	1.007***	0.802***	0.807***	0.801***	1.108***
	(0.189)	(0.215)	(0.192)	(0.299)	(0.196)	(0.203)	(0.199)	(0.281)
Constant	8.455***	8.341***	8.299***	7.492***	8.379***	7.651***	8.384***	8.032***
	(1.229)	(3.006)	(1.164)	(1.273)	(1.317)	(2.708)	(1.316)	(1.320)
Observations	306	306	306	255	306	306	306	255
Nb. states	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.991	0.993	0.993	0.993	0.991
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (2.3) and includes all fixed effects. We supplement our benchmark specifications reported in col. 1 and 5 we alternatives definition of the epidemiological term ($MY^S_{r,t}$). We compute $MY^S_{r,t}$ (Const. $\widehat{k}^S_{i,r,t}$) by keeping the immigration shares constant, at their 1960-2010 average levels. We $MY^S_{r,t}$ (Const. $y_{i,t}$) by keeping the levels of GDP per capita at origin ($log(y_{i,t})$) constant, at their 1960-2010 average level. We compute $MY^S_{r,t}$ ($y_{i,Entry}$) combining annual data on GDP per capita at origin with individual data on the year of arrival in the US. Each immigration share is multiplied by the average level of GDP per capita prevailing in the year of immigration to the US which allows us to capture the norms and values that immigrants bring with them when they migrate. Due to data limitations, this variable cannot be computed for the year 1960. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

Figure A12: Diversity among immigrants ($MD^A_{r,t}$) in the US states



Source: Authors' elaboration on IPUMS-US data. Diversity among immigrants is defined as in Eq. (2.2).

Table A22: First-Stage regressions High-skilled (Alternative IV Strategies)

	(1) $MD^H_{r,t}$	(2) $MD^H_{r,t}$	(3) $MD^H_{r,t}$	(3) $m^H_{r,t}$	(4) $MY^H_{r,t}$	(4) $m^H_{r,t}$	(5) $MD^H_{r,t}$	(5) $MY^H_{r,t}$	(5) $m^H_{r,t}$
$MD^H_{r,t}$(Shift-share)	0.751***								
	(0.036)								
$MD^H_{r,t}$(Feyrer)		0.304***	0.302***	-0.003			0.318***	-0.236*	-0.002
		(0.030)	(0.032)	(0.013)			(0.027)	(0.093)	(0.014)
$m^H_{r,t}$ (Feyrer)			-0.921***	1.600***	-0.189	1.483***	-0.678***	0.513	1.528***
			(0.188)	(0.141)	(0.660)	(0.141)	(0.155)	(0.709)	(0.143)
$MY^H_{r,t}$ (Feyrer)					0.349***	-0.013***	0.031***	0.323***	-0.014***
					(0.020)	(0.004)	(0.008)	(0.020)	(0.004)
$MD^H_{r,t}$					-0.919***	-0.043			
					(0.237)	(0.027)			
$m^H_{r,t}$	-0.266	-0.524***							
	(0.137)	(0.108)							
$MY^H_{r,t}$			0.007	-0.020*					
			(0.021)	(0.009)					
$log(Pop_{r,t})$	-0.004	0.009	0.020	-0.016	0.002	-0.012	0.010	-0.011	-0.013
	(0.012)	(0.015)	(0.017)	(0.010)	(0.035)	(0.009)	(0.016)	(0.039)	(0.010)
$log(Urb_{r,t})$	0.005	-0.027	-0.007	-0.035*	0.058	-0.035*	-0.008	0.059	-0.036*
	(0.022)	(0.078)	(0.081)	(0.017)	(0.068)	(0.014)	(0.075)	(0.071)	(0.015)
$log(Hum_{r,t})$	0.067	0.081	0.148	-0.119	-0.044	-0.103	0.130	-0.175	-0.111
	(0.047)	(0.072)	(0.084)	(0.062)	(0.171)	(0.055)	(0.075)	(0.204)	(0.058)
Observations	306	306	306	306	306	306	306	306	306
R-squared	0.717	0.660	0.642	0.853	0.748	0.862	0.680	0.719	0.860
Angrist-Pischke F-test	439.93	103.94	87.26	99.19	233.71	108.93	138.18	237.69	95.81

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Source: Authors' elaboration on IPUMS-US data. This table provides the results of the first-stage regressions under different IV (Shift share and Feyrer).

Table A23: First-Stage regressions Low-skilled (Alternative IV Strategies)

	(1) $MD^L_{r,t}$	(2) $MD^L_{r,t}$	(3) $MD^L_{r,t}$	(3) $m^L_{r,t}$	(4) $MY^L_{r,t}$	(4) $m^L_{r,t}$	(5) $MD^L_{r,t}$	(5) $MY^L_{r,t}$	(5) $m^L_{r,t}$
$MD^L_{r,t}$ (Shift-Share)	0.788***								
	(0.083)								
$MD^L_{r,t}$ (Feyrer)		0.416***	0.390***	0.022			0.393***	-0.177**	0.014
		(0.050)	(0.048)	(0.016)			(0.047)	(0.058)	(0.013)
$m^L_{r,t}$ (Feyrer)			-1.166***	1.379***	-0.592	1.337***	-1.150***	-0.608	1.346***
			(0.304)	(0.129)	(0.428)	(0.120)	(0.300)	(0.446)	(0.119)
$MY^L_{r,t}$ (Feyrer)					0.385***	-0.006	0.008	0.371***	-0.005
					(0.027)	(0.004)	(0.013)	(0.027)	(0.004)
$MD^L_{r,t}$					-0.213	0.014			
					(0.109)	(0.022)			
$m^L_{r,t}$	-0.863***	-0.603*							
	(0.229)	(0.232)							
$MY^L_{r,t}$			0.006	0.010					
			(0.039)	(0.014)					
$log(Pop_{r,t})$	-0.024	-0.052	-0.028	-0.008	0.180**	0.000	-0.029	0.219***	-0.003
	(0.048)	(0.040)	(0.041)	(0.013)	(0.059)	(0.010)	(0.039)	(0.061)	(0.012)
$log(Urb_{r,t})$	-0.043	-0.206	-0.164	-0.049*	-0.151	-0.050*	-0.166	-0.137	-0.050*
	(0.091)	(0.141)	(0.138)	(0.021)	(0.111)	(0.022)	(0.138)	(0.100)	(0.022)
$log(Hum_{r,t})$	0.639**	0.232	0.311	-0.145	-0.114	-0.133*	0.306	-0.070	-0.139
	(0.192)	(0.215)	(0.213)	(0.079)	(0.272)	(0.064)	(0.212)	(0.310)	(0.077)
Observations	306	306	306	306	306	306	306	306	306
R-squared	0.519	0.604	0.617	0.881	0.762	0.880	0.618	0.769	0.881
Angrist-Pischke F-test	91.14	70.12	73.14	75.92	198.22	120.62	76.08	183.86	85.90

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Source: Authors' elaboration on IPUMS-US data. This table provides the results of the first-stage regressions under different IV (Shift share and Feyrer).

Table A24: Robustness of FE regressions to age of entry, 1970-2010.

	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$	$S=H$
Age of entry >	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$MD^H_{r,t}$	0.733**	0.739**	0.703**	0.676**	0.694**	0.655**	0.645**	0.638**	0.564**	0.550**	0.548**	0.535**	0.545**	0.526**	0.414**	0.218***	0.210***	0.226**	0.244**	0.231**	0.216**
	(0.308)	(0.310)	(0.307)	(0.305)	(0.326)	(0.322)	(0.307)	(0.308)	(0.222)	(0.225)	(0.222)	(0.218)	(0.222)	(0.215)	(0.160)	(0.079)	(0.078)	(0.090)	(0.103)	(0.094)	(0.087)
$m^H_{s,t}$	0.633	0.641	0.647	0.652	0.677	0.659	0.681	0.690	0.699	0.695	0.714	0.736	0.735	0.771	0.752	0.660	0.665	0.682	0.745	0.762	0.765
	(0.385)	(0.395)	(0.408)	(0.414)	(0.433)	(0.433)	(0.445)	(0.449)	(0.441)	(0.444)	(0.454)	(0.462)	(0.470)	(0.494)	(0.498)	(0.487)	(0.522)	(0.547)	(0.591)	(0.649)	(0.733)
$ln(Population_{s,t})$	-0.087	-0.090	-0.090	-0.092	-0.094	-0.091	-0.092	-0.093	-0.086	-0.085	-0.085	-0.085	-0.085	-0.086	-0.080	-0.068	-0.067	-0.066	-0.066	-0.066	-0.065
	(0.066)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.068)	(0.068)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.068)	(0.068)	(0.068)	(0.068)	(0.068)	(0.067)	(0.067)	(0.067)
$ln(Urban_{s,t})$	0.226	0.229	0.218	0.225	0.216	0.218	0.217	0.219	0.204	0.205	0.215	0.221	0.213	0.211	0.207	0.190	0.196	0.190	0.183	0.184	0.181
	(0.154)	(0.154)	(0.152)	(0.154)	(0.152)	(0.154)	(0.150)	(0.151)	(0.148)	(0.149)	(0.149)	(0.150)	(0.149)	(0.148)	(0.149)	(0.157)	(0.157)	(0.157)	(0.155)	(0.158)	(0.158)
$ln(College_{s,t})$	0.887***	0.881***	0.906***	0.910***	0.915***	0.921***	0.924***	0.924***	0.951***	0.956***	0.948***	0.940***	0.946***	0.958***	0.970***	1.001***	0.996***	0.997***	1.000***	0.999***	1.001***
	(0.278)	(0.278)	(0.283)	(0.283)	(0.286)	(0.286)	(0.288)	(0.288)	(0.289)	(0.289)	(0.289)	(0.288)	(0.289)	(0.293)	(0.292)	(0.291)	(0.289)	(0.292)	(0.296)	(0.295)	(0.297)
Constant	6.589***	6.618***	6.661***	6.673***	6.721***	6.702***	6.724***	6.726***	6.712***	6.704***	6.682***	6.681***	6.691***	6.706***	6.732***	6.749***	6.726***	6.727***	6.743***	6.751***	6.756***
	(1.315)	(1.313)	(1.295)	(1.294)	(1.281)	(1.288)	(1.271)	(1.269)	(1.272)	(1.276)	(1.280)	(1.283)	(1.280)	(1.268)	(1.270)	(1.280)	(1.284)	(1.284)	(1.275)	(1.283)	(1.285)
R-squared	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.990	0.990	0.990	0.991	0.990	0.990
	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$	$S=L$
Age of entry >	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$MD^L_{r,t}$	0.062	0.059	0.060	0.057	0.044	0.043	0.039	0.037	0.037	0.029	0.027	0.029	0.022	0.018	0.020	0.027	0.027	0.022	0.022	0.030	0.026
	(0.071)	(0.071)	(0.072)	(0.071)	(0.073)	(0.072)	(0.071)	(0.072)	(0.071)	(0.073)	(0.073)	(0.075)	(0.077)	(0.077)	(0.077)	(0.077)	(0.078)	(0.078)	(0.075)	(0.080)	(0.082)
$m^L_{s,t}$	0.462*	0.459*	0.457*	0.458*	0.443	0.444	0.437	0.441	0.452	0.444	0.439	0.457	0.470	0.482	0.520	0.534	0.570	0.601	0.637	0.683	0.733
	(0.258)	(0.258)	(0.259)	(0.262)	(0.268)	(0.270)	(0.272)	(0.275)	(0.280)	(0.284)	(0.288)	(0.300)	(0.313)	(0.330)	(0.353)	(0.379)	(0.409)	(0.443)	(0.479)	(0.518)	(0.570)
$ln(Population_{s,t})$	-0.085	-0.085	-0.084	-0.083	-0.083	-0.082	-0.081	-0.081	-0.081	-0.080	-0.079	-0.079	-0.079	-0.078	-0.078	-0.077	-0.077	-0.076	-0.076	-0.076	-0.075
	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)
$ln(Urban_{s,t})$	0.242	0.242	0.241	0.240	0.233	0.233	0.232	0.232	0.232	0.229	0.228	0.228	0.226	0.225	0.226	0.228	0.229	0.228	0.229	0.230	0.231
	(0.174)	(0.174)	(0.174)	(0.174)	(0.175)	(0.175)	(0.175)	(0.175)	(0.175)	(0.176)	(0.176)	(0.176)	(0.177)	(0.177)	(0.176)	(0.176)	(0.178)	(0.178)	(0.176)	(0.175)	(0.175)
$ln(College_{s,t})$	1.038***	1.037***	1.036***	1.038***	1.049***	1.047***	1.049***	1.052***	1.053***	1.057***	1.055***	1.051***	1.054***	1.051***	1.045***	1.027***	1.023***	1.025***	1.019***	1.010***	1.010***
	(0.307)	(0.308)	(0.310)	(0.309)	(0.312)	(0.311)	(0.311)	(0.312)	(0.311)	(0.313)	(0.312)	(0.312)	(0.313)	(0.311)	(0.310)	(0.307)	(0.306)	(0.307)	(0.306)	(0.308)	(0.309)
Constant	6.858***	6.857***	6.850***	6.848***	6.857***	6.853***	6.848***	6.847***	6.841***	6.843***	6.840***	6.841***	6.850***	6.851***	6.849***	6.847***	6.850***	6.852***	6.851***	6.854***	6.845***
	(1.278)	(1.277)	(1.277)	(1.278)	(1.280)	(1.281)	(1.283)	(1.283)	(1.283)	(1.284)	(1.287)	(1.288)	(1.290)	(1.292)	(1.291)	(1.294)	(1.294)	(1.295)	(1.295)	(1.295)	(1.293)
R-squared	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990
Observations	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
Nb. states	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on IPUMS-US data. Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the state level.

Table A25: System GMM. Internal instruments.
Results by skill group (Dep= $log(y_{r,t})$)

	(1) $S=H$	(2) $S=H$	(3) $S=H$	(4) $S=H$	(5) $S=H$	(6) $S=L$	(7) $S=L$	(8) $S=L$	(9) $S=L$	(10) $S=L$
$log(y_{r,t-10})$	0.546***	0.269	0.307*	0.314*	0.309*	0.555***	0.293*	0.338*	0.339*	0.370*
	(0.116)	(0.190)	(0.179)	(0.168)	(0.174)	(0.189)	(0.172)	(0.188)	(0.185)	(0.188)
$MD^S_{r,t}$	1.792***	1.295**	1.173**	1.088**	1.231***	0.238*	0.371***	0.358***	0.362***	0.385***
	(0.527)	(0.580)	(0.467)	(0.448)	(0.374)	(0.119)	(0.103)	(0.102)	(0.103)	(0.104)
$m^S_{r,t}$	1.325***	0.892	0.896*	0.862	0.953*	0.313	0.147	0.357	0.541*	0.574*
	(0.483)	(0.563)	(0.485)	(0.519)	(0.509)	(0.324)	(0.363)	(0.401)	(0.318)	(0.321)
$log(Pop_{r,t})$	-0.065	-0.101	-0.082	-0.082	-0.083	-0.039	-0.093*	-0.067	-0.074	-0.062
	(0.046)	(0.072)	(0.061)	(0.051)	(0.059)	(0.044)	(0.053)	(0.054)	(0.049)	(0.050)
$log(Urb_{r,t})$	0.017	0.416	0.304	0.323	0.306	0.204	0.593**	0.397	0.358	0.280
	(0.125)	(0.431)	(0.316)	(0.309)	(0.255)	(0.145)	(0.277)	(0.266)	(0.258)	(0.203)
$log(Hum_{r,t})$	0.563**	0.356	0.439	0.395	0.465	0.359	-0.148	0.095	0.315	0.302
	(0.260)	(0.436)	(0.344)	(0.321)	(0.334)	(0.266)	(0.442)	(0.416)	(0.287)	(0.324)
Constant	2.979**	5.498***	5.261***	5.280***	5.133***	2.888*	4.895***	4.567***	4.445***	4.364**
	(1.184)	(1.897)	(1.794)	(1.514)	(1.748)	(1.583)	(1.173)	(1.365)	(1.184)	(1.650)
Observations	255	255	255	255	255	255	255	255	255	255
Nb. states	51	51	51	51	51	51	51	51	51	51
Nb. instruments	56	26	31	36	41	56	26	31	36	41
Nb. lags (endogenous var.)	2	2	3	4	5	2	2	3	4	5
Collapsed matrix	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.006	0.021	0.020	0.015	0.019	0.006	0.012	0.011	0.012	0.010
AR(2)	0.134	0.118	0.121	0.126	0.124	0.256	0.200	0.210	0.215	0.222
Hansen J (p-value)	0.554	0.056	0.158	0.218	0.304	0.505	0.202	0.168	0.244	0.170

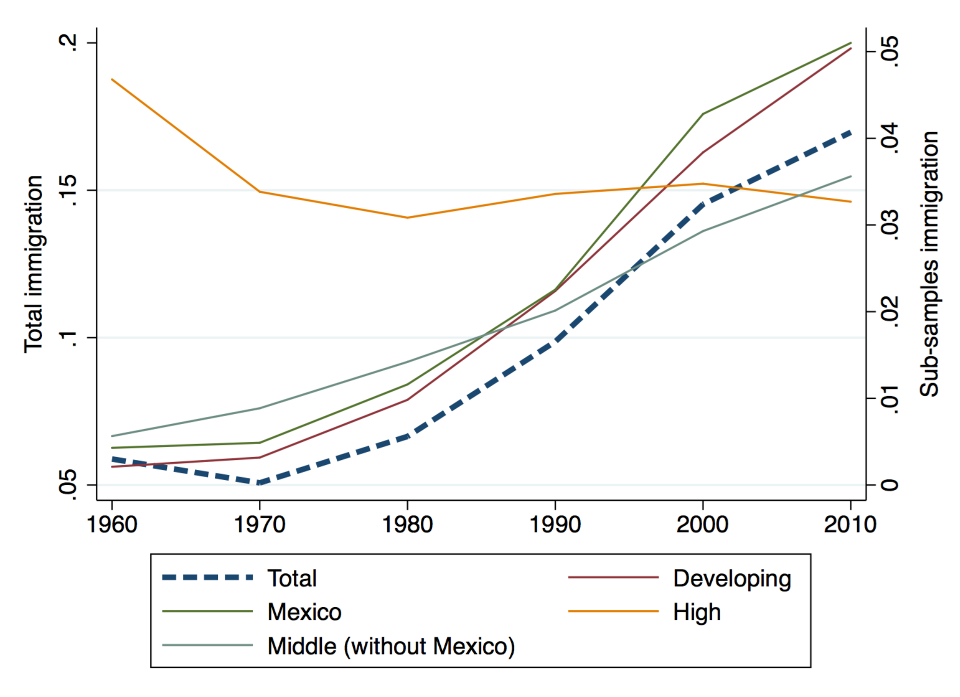
Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors in parentheses are clustered at the state level. Source: Authors' elaboration on IPUMS-US data. The lagged dependent variable is always treated as predetermined and instrumented with its own first to second lags. $MD^S_{r,t}$; $m^S_{r,t}$; $log(Pop_{r,t})$; $log(Urb_{r,t})$ and $log(Hum_{r,t})$ are treated as endogenous variables and instrumented with their own first to X lags. The number of lags X is reported in the table. From columns (2) to (5) and (7) to (10) the matrix of endogenous variable is collapsed in order to keep the number of instruments below the number of states.

Table A26: System GMM. External instruments.
Results by skill group (Dep= $log(y_{r,t})$)

	(1) $S=H$	(2) $S=H$	(3) $S=H$	(4) $S=H$	(5) $S=L$	(6) $S=L$	(7) $S=L$	(8) $S=L$
$log(y_{r,t-10})$	0.184	0.215	0.202	0.228	0.293	0.361*	0.308*	0.357**
	(0.166)	(0.144)	(0.169)	(0.153)	(0.197)	(0.201)	(0.174)	(0.171)
$MD^S_{r,t}$	1.504**	1.587***	1.675***	1.791***	0.254*	0.263*	0.435***	0.446***
	(0.686)	(0.521)	(0.466)	(0.486)	(0.131)	(0.146)	(0.116)	(0.119)
$m^S_{r,t}$	1.353*	1.543**	1.435**	1.660**	0.266	0.546	0.416	0.655
	(0.684)	(0.618)	(0.603)	(0.620)	(0.388)	(0.416)	(0.433)	(0.450)
$log(Pop_{r,t})$	-0.132**	-0.119***	-0.128*	-0.122**	-0.088	-0.069	-0.092*	-0.080*
	(0.059)	(0.044)	(0.064)	(0.052)	(0.067)	(0.058)	(0.048)	(0.042)
$log(Urb_{r,t})$	0.382	0.267	0.336	0.242	0.474	0.280	0.472*	0.348
	(0.416)	(0.298)	(0.367)	(0.291)	(0.339)	(0.300)	(0.264)	(0.237)
$log(Hum_{r,t})$	0.349	0.503	0.332	0.502	0.029	0.280	-0.215	0.054
	(0.476)	(0.407)	(0.440)	(0.392)	(0.325)	(0.263)	(0.371)	(0.316)
Constant	6.729***	6.336***	6.545***	6.143***	5.115***	4.672***	5.320***	4.811***
	(1.775)	(1.435)	(1.667)	(1.323)	(1.347)	(1.456)	(1.333)	(1.306)
Observations	255	255	255	255	255	255	255	255
Nb. states	51	51	51	51	51	51	51	51
Nb. instruments	25	33	25	33	25	33	25	33
Nb. lags (endogenous var.)	2	4	2	4	2	4	2	4
Collapsed matrix	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.032	0.023	0.036	0.029	0.016	0.014	0.014	0.008
AR(2)	0.097	0.104	0.097	0.102	0.190	0.206	0.188	0.213
Hansen J (p-value)	0.189	0.306	0.202	0.272	0.145	0.189	0.232	0.303

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Source: Authors' elaboration on IPUMS-US data. The lagged dependent variable is always treated as predetermined and instrumented with its own first to second lags. $m^S_{r,t}$; $log(Pop_{r,t})$; $log(Urb_{r,t})$ and $log(Hum_{r,t})$ are treated as endogenous variables and instrumented with their own first to X lags. The number of lags X is reported in the table. The matrix of endogenous variables is collapsed in order to keep the number of instruments below the number of states. In columns (1), (2), (5) and (6), $MD^S_{r,t}$ is instrumented using the augmented shift-share strategy while in columns (3), (4), (7) and (8), $MD^S_{r,t}$ is instrumented using predictions of the gravity-like strategy *a la* Feyrer (2009).

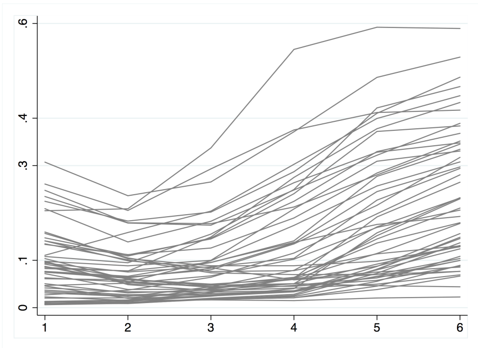
Figure A13: US immigration rate, 1960-2010 (as percentage of total population)



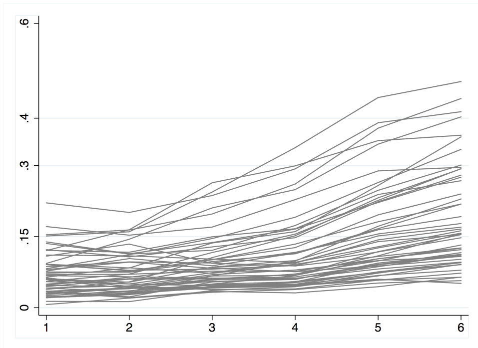
Source: Authors' elaboration on IPUMS-US data.

Notes: The "Total immigration rate" is defined as the ratio of the total stock of foreign-born individuals to the total population of the destination country or region. The "Immigration rate from high, developing and middle income countries" is defined as the ratio of the total stock of foreign-born individuals originating respectively from high, developing and middle countries to the total population of the destination country or region. The definition of a high, developing and middle income countries follows the World Bank classification of 2015.

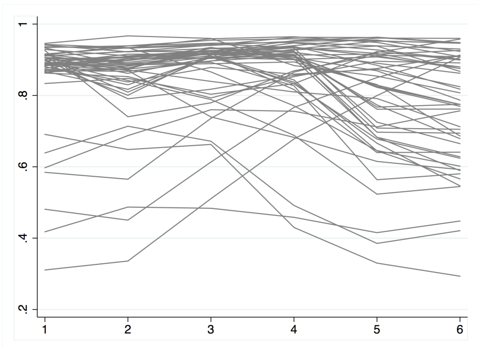
Figure A14: Global trends in birtplace diversity in the US states



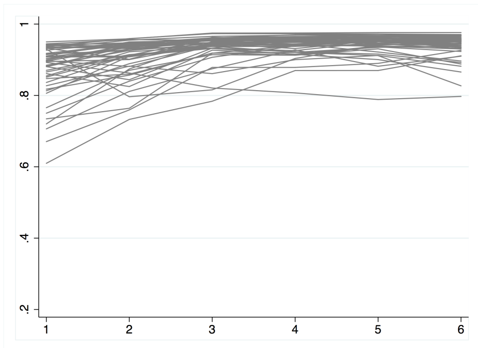
$TD^{L}_{r,t}$



$TD^{H}_{r,t}$



$MD^{L}_{r,t}$



$MD^{H}_{r,t}$

Source: Authors' elaboration on IPUMS-US data. Diversity among residents is defined as in Eq. (1) in the paper. Diversity among immigrants is defined as in Eq (2) in the paper.

Empirical analysis for the OECD countries

This appendix presents the results of a complementary analysis conducted on the 34 member states of the OECD. Despite the drawbacks of the OECD data (e.g. absence of information on the skill composition of migration stocks), we only want to verify whether the results obtained for the US states are not invalidated when using cross-country data, as the US is usually considered as one of the most attractive countries for (high-skilled) migrants.

Population data at the country level for the OECD member states are available from the Global Migrant Stock database described in Özden*et al* (2011). This database documents the bilateral stocks and shares of international migrants, $k^A_{i,r,t}$, in the population of each OECD country r, by country of origin i, and by year t. Özden*et al* (2011) collected and harmonized over 1,000 censuses and population registers to construct comprehensive matrices of origin-destination stocks that correspond to the last five completed census rounds, i.e. for the period 1960-2000 in 10-year intervals. They specified a standard and common set of countries for the entire period, disaggregating data for the countries that no longer exist on the basis of more recent migration figures. There is no artificial variation due to the dislocation of the Eastern Block. We expanded the database by adding the share of native citizens, $k^A_{r,r,t}$, in order to match the total population data. We also added the year 2010 using the bilateral stock estimates of the United nations (2013).[29] Hence, our OECD database covers the same period (1960-2010) in ten-year intervals as the IPUMS data; results for the US states and for OECD countries are then comparable. Still, compared to the IPUMS data, the OECD data suffer from two drawbacks. The first drawback is that it does not report the educational structure of migration stocks. The second drawback is that many imputations were used to fill the missing bilateral cells.

We compute our indices of birthplace diversity, $TD^A_{r,t}$ and $MD^A_{r,t}$, for each OECD member state. Figure A19 shows that most OECD member states have experienced increasing immigration rates and total diversity indices ($TD^A_{r,t}$) in the aftermath of WW2. This is particularly the case after the year 1980.[30]

[29] The list of the 34 OECD member states as well as descriptive statistics are available in Table A27.

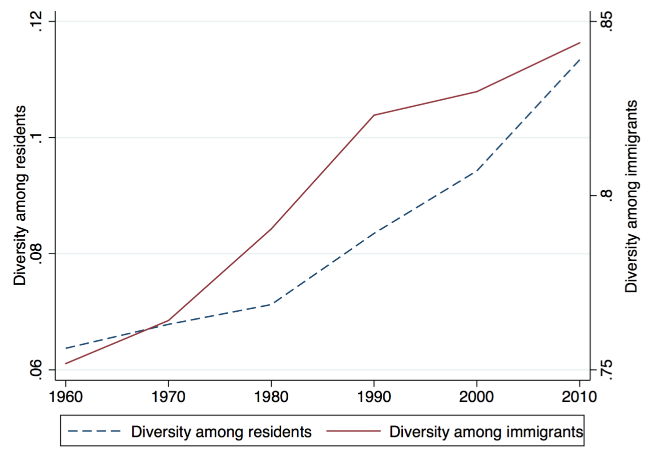
[30] Diversity trends for OECD countries are described in Figure A20.

Compared to the US, the average level of diversity among immigrants ($MD^A_{r,t}$) increased more strongly. Unsurprisingly, this average trend conceals important disparities across countries. Although a rise in the variety of immigrants was observed in most countries, diversity decreased in countries such as the US, Mexico or Slovakia.

Results of the OECD regressions are depicted in Table A28. The first three columns of the table show the effect of birthplace diversity among immigrants on GDP per capita, regardless of the educational structure. It is worth noticing that these estimates include the log ratio of trade to GDP, the logarithm of the population, the log of the number of years of schooling in the working-age population and the *Polity2* index of democracy as covariates. As in the US state sample, the effect of birthplace diversity on GDP per capita is strongly positive and significant at the 1% level when using pooled OLS, OLS-FE and the gravity-like IV strategy *a la* Feyrer (2009)). The magnitude of the effect is larger than in the US sample: a one standard deviation change in birthplace diversity is associated with a 13% increase in GDP per capita. This implies that the Japanese level of GDP per capita would be 1,770 dollars greater if Japan had the same diversity index as the US in 2010.

In the remainder of the table, we combine the data of Özden*et al* (2011) with cross-sectional data on the skill structure of migration stocks. We use the database of Artuc *et al* (2015), which documents the proportion on college-educated immigrants in all OECD countries for the years 1990 and 2000. In col. 4, we add an interaction term, the product of the average birthplace diversity index by the proportion of college-educated immigrants observed in 2000. Despite collinearity with the non-interacted index, the interaction term is positive and significant, whereas the average index of diversity looses significance. This suggests that the effect of birthplace diversity increases with the educational level of migrants. At the median level of the proportion of college graduates (22%), the marginal effect of birthplace diversity on GDP per capita is significant at the 1% level, and the coefficient is almost equal to that obtained with the US state sample (0.506). The results of col. 4 are illustrated on Figure 4.3. Interestingly, this figure shows that the effect of birthplace diversity is insignificant when the bilateral migration stocks is mainly composed of low-skilled migrants. However, above 20% of college graduates, the effect of diversity

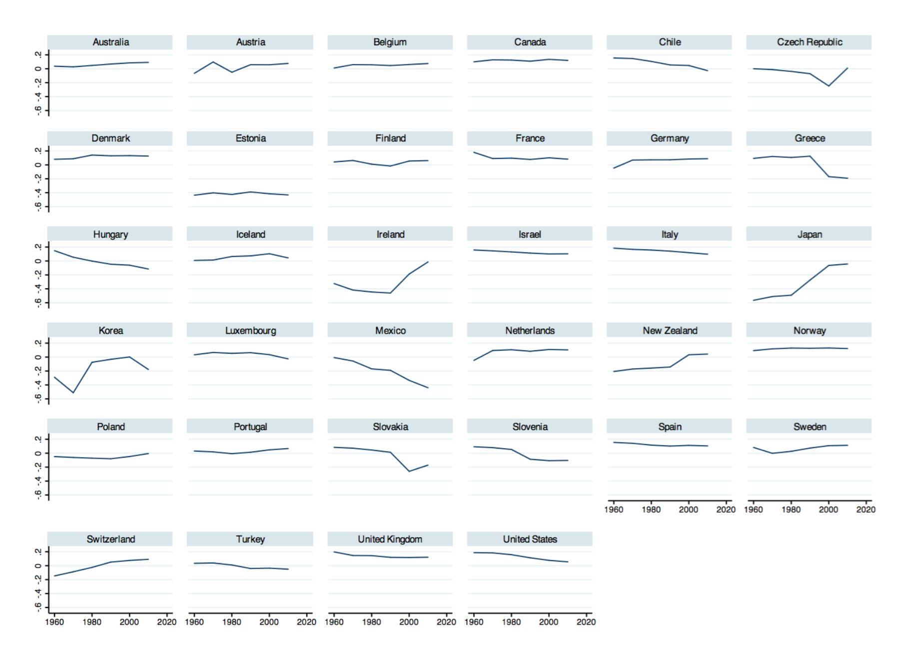
Figure A19: Trends in birthplace diversity in the OECD member states, 1960-2010



Notes: Diversity among residents is defined as in Eq. (2.1), whereas diversity among immigrants is defined as in Eq. (2.2). Source: Authors' elaboration on Özden *et al.* (2011).

becomes significant, and increases with the proportion of college graduates. Again, this suggests that the effect of birthplace diversity on macroeconomic performance is skill-specific. In the two last columns of Table A28, we tentatively proxy the stocks of high-skilled and low-skilled migrants using the recent shares of college-educated migrants provided in Artuc *et al* (2015). We use the 1990 skill shares to split the bilateral migration stocks observed in 1960, 1970, 1980 and 1990; we use the 2000 skill shares to split the stocks observed in 2000 and 2010. We find a positive and significant effect of birthplace diversity in both regressions, and a greater effect for high-skilled diversity. Although we should not give too much credit to these results, we confirm that the patterns obtained for the US states are not invalidated when using cross-country data.

Figure A20: Diversity among immigrants ($MD^A_{r,t}$) in the OECD countries



Source: Authors' elaboration on Özden*et al* (2011) data.

Source: Authors' elaboration on Özden*et al* (2011). Diversity among residents is defined as in Eq. (1). Diversity among immigrants is defined as in Eq. (2).

Table A27: Descriptives statistics for the OECD countries (34)

	$log(y_{r,t})$	$MD^A_{r,t}$	$m^A_{r,t}$	$log(Pop_{r,t})$	$log(Trade_{r,t})$	$Democ_{r,t}$	$log(Hum_{r,t})$
Australia	9.655	0.860	0.202	16.563	3.486	1.000	2.338
Austria	9.543	0.830	0.113	15.855	4.226	1.000	2.095
Belgium	9.576	0.852	0.081	16.114	4.699	0.983	2.151
Canada	9.695	0.920	0.168	17.053	3.948	1.000	2.298
Chile	8.843	0.881	0.012	16.294	3.851	0.758	1.958
Czech Republic	9.018	0.741	0.035	16.132	4.349	0.542	2.366
Denmark	9.690	0.917	0.044	15.445	4.266	1.000	2.046
Estonia	9.254	0.384	0.194	14.134	4.739	0.475	2.201
Finland	9.516	0.836	0.018	15.401	4.047	1.000	1.941
France	9.594	0.905	0.099	17.816	3.708	0.900	1.901
Germany	9.544	0.857	0.109	18.194	3.814	0.763	2.193
Greece	9.034	0.814	0.038	16.096	3.643	0.792	2.024
Hungary	8.713	0.797	0.038	16.146	4.575	0.575	2.213
Iceland	9.561	0.852	0.045	12.384	4.361	1.000	2.064
Ireland	9.290	0.492	0.071	15.057	4.685	1.000	2.186
Israel	9.286	0.926	0.395	15.217	4.169	0.975	2.305
Italy	9.465	0.945	0.035	17.831	3.646	1.000	1.898
Japan	9.465	0.476	0.010	18.550	3.121	1.000	2.191
Korea	8.607	0.620	0.003	17.443	3.881	0.708	1.965
Luxembourg	9.896	0.838	0.248	12.863	5.309	1.000	2.029
Mexico	8.628	0.601	0.005	18.107	3.459	0.508	1.512
Netherlands	9.647	0.875	0.061	16.468	4.682	1.000	2.220
New Zealand	9.492	0.700	0.162	14.998	4.047	1.000	2.396
Norway	9.655	0.921	0.049	15.244	4.299	1.000	2.256
Poland	8.632	0.748	0.043	17.379	3.973	0.533	2.114
Portugal	8.999	0.829	0.038	16.083	3.978	0.675	1.405
Slovakia	8.911	0.764	0.011	15.412	4.335	0.550	2.340
Slovenia	9.221	0.789	0.054	14.429	4.606	0.467	2.228
Spain	9.142	0.923	0.040	17.440	3.510	0.708	1.820
Sweden	9.666	0.868	0.088	15.946	4.101	1.000	2.263
Switzerland	9.855	0.794	0.197	15.691	4.398	1.000	2.352
Turkey	8.402	0.793	0.024	17.659	3.005	0.775	1.210
United Kingdom	9.587	0.942	0.066	17.859	3.885	1.000	2.162
United States	9.906	0.929	0.085	19.307	2.855	1.000	2.449

Source: Authors' elaboration on Özden *et al* (2011) .

Table A28: Birthplace diversity in a cross-country setting. Regressions for the OECD member states (Dep= $log(y_{r,t})$)

	(1) OLS $S=A$	(2) FE $S=A$	(3) 2SLS $S=A$	(4) 2SLS $S=A$	(5) 2SLS $S=H$	(6) 2SLS $S=L$
$MD^S_{r,t}$	0.834***	0.788**	0.940***	-0.318	1.016***	0.901***
	(0.274)	(0.292)	(0.245)	(0.500)	(0.316)	(0.259)
$MD^A_{r,t} \times CollMig_{r,t}$				3.754**		
				(1.839)		
$m^S_{r,t}$	0.683	-0.291	0.148	0.020	0.107	-0.096
	(0.911)	(0.333)	(0.429)	(0.374)	(0.070)	(0.172)
$log(Pop_{r,t})$	-0.002	0.005	0.076	-0.001	0.073	0.025
	(0.027)	(0.144)	(0.143)	(0.138)	(0.145)	(0.152)
$log(Trade_{r,t})$	0.073	0.150	0.157*	0.146*	0.140	0.160*
	(0.099)	(0.089)	(0.086)	(0.080)	(0.092)	(0.083)
$log(Hum_{r,t})$	0.507***	0.370*	0.363**	0.425**	0.425**	0.347**
	(0.152)	(0.187)	(0.181)	(0.181)	(0.196)	(0.170)
$Democ_{r,t}$	0.578***	0.039	0.076	0.046	0.039	0.068
	(0.105)	(0.078)	(0.070)	(0.071)	(0.068)	(0.077)
Constant	6.420***	6.750**				
	(0.702)	(2.625)				
Total effect of $MD^A_{r,t}$				0.506***		
				(0.192)		
Observations	204	204	204	204	204	204
Nb. countries	34	34	34	34	34	34
R-squared	0.750	0.901	0.899	0.908	0.897	0.903
Time fixed effects	No	Yes	Yes	Yes	Yes	Yes
Country fixed effects	No	Yes	Yes	Yes	Yes	Yes
K-P F-$Test^\dagger$			40.89	92.66	31.64	35.18
Stock Yogo			7.03/4.58	13.43/8.18	7.03/4.58	7.03/4.58

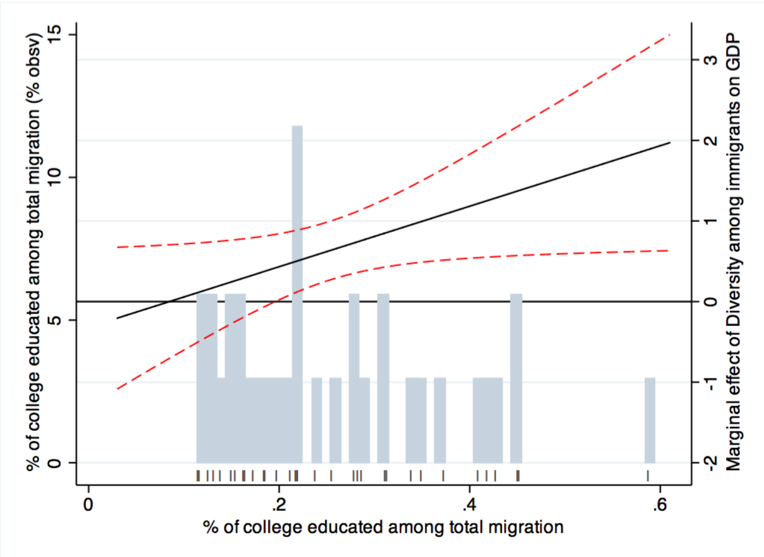
Notes. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the country level. The $CollMig_{r,t}$ variable is the share of college-educated migrants in the total stock of immigrants of the receiving country in 2000. The total effect of diversity in col. (4) is computed for $CollMig_{r,t}$ at its median level. Diversity and immigration rates are instrumented using the gravity-like IV strategy *a la* Feyrer (2009). Zero-stage estimates are available in the appendix (Table A20). †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). The sample includes the 34 OECD member states from 1960 to 2010. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of trade ($log(Trade_{r,t})$), the log of the number of years of schooling in the working-age population ($log(Hum_{r,t})$), and the Polity2 index of democracy ($Democ_{r,t}$).

Table A29: Zero-stage estimates (PPML): gravity model *a la* Feyrer (2009)

	(1) All OECD $log(Stock_{i,c,t})$
$log(Dist_{i,r}) \times I_{1960}$	-0.816***
	(0.155)
$log(Dist_{i,r}) \times I_{1970}$	-0.876***
	(0.151)
$log(Dist_{i,r}) \times I_{1980}$	-0.774***
	(0.139)
$log(Dist_{i,r}) \times I_{1990}$	-0.683***
	(0.131)
$log(Dist_{i,r}) \times I_{2000}$	-0.617***
	(0.125)
$log(Dist_{i,r}) \times I_{2010}$	-0.573***
	(0.124)
$Bord_{i,r}$	0.755***
	(0.219)
$Lang_{i,r}$	1.565***
	(0.218)
Constant	15.078***
	(1.322)
Observations	38964
Nb. origin	191
Nb. destination	34
R-squared	0.727
Year dummies	Yes
Origin dummies	Yes
Destination dummies	Yes

Notes. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the state/country-pair level. Distances data are not available for Liechtenstein, Luxembourg and Holy See.

Figure A21: Marginal effect of diversity conditional to the share of college graduates in the 2000 immigrant stock



Notes: The solid line is based on col. 4 of Table A28. It represents the marginal effect of birthplace diversity on GDP per capita conditional on the percentage of college graduates in the total stock of bilateral migrants in 2000. The histogram indicates the percentage of observations of the modifying variable, and each mark on the vertical axis represents one OECD country. The dashed line depicts the upper and lower bounds of the 95% confidence interval.

Augmented diversity index

In this appendix, we report the results obtained when the standard index of birthplace diversity is replaced by an augmented index that accounts for the genetic distance between the countries of origin of immigrants. Indeed, the birthplace diversity index $MD^S_{r,t}$ does not account for the cultural distance between origin and destination countries. It assumes that all groups are culturally equidistant from each other. Another extension consists therefore in multiplying the probability that two randomly-drawn immigrants were born in two different countries by a measure of cultural distance between these two countries. For the latter, we use the database on genetic distance between countries, constructed by Spolaore and Wacziarg (2015). Genetic distance is based on blood sample and proxies the time since two populations had common ancestors. Spolaore and Wacziarg (2015) find a pattern of positive and significant relationships between genetic distance and various measures of cultural distance, including language, religion, values, and norms. We thus use the augmented index of cultural diversity to investigate whether the variety effect is associated with the genetic distance between countries of origin. The augmented version of the diversity index, computed for the immigrant population, is defined as:

$$MD^{S,G}_{r,t} = \sum_{i \neq r}^{I} \widehat{k}^S_{i,r,t} \sum_{j \neq i,r}^{J} \widehat{k}^S_{j,r,t} d^G_{i,j}, \tag{2.8}$$

where $d^G_{i,j} \in [0,1]$ is a normalized genetic distance between population from country i and country j. The correlation between the augmented and the unweighted diversity indices is equal to 0.69 for the college-educated population, and to 0.38 for the less educated. The OLS-FE estimates point in the same direction as our benchmark regressions. The magnitude and the significance of the estimates are not statistically different from the ones reported in the previous tables. Using the gravity-like IV strategy, the effect of high-skilled diversity remains positive and significant. We lose significance under the augmented shift-share IV strategy. However, the shift-share statistics are less convincing when we introduce more complexity in the diversity index. As far as low-skilled diversity is concerned, the effect remains insignificant in OLS-FE regressions and also in IV estimates. Overall, accounting for cultural distance between countries does not add that much to our analysis.

Table A30: Robustness to the measure of diversity.
Accounting for genetic distance between countries (Dep= $log(y_{r,t})$)

	(1) FE OLS $S = H$	(2) 2SLS Shift-Share $S = H$	(3) 2SLS Feyrer $S = H$	(4) FE OLS $S = L$	(5) 2SLS Shift-Share $S = L$	(6) 2SLS Feyrer $S = L$
$MD^{S,G}_{r,t}$	0.658***	0.558	1.092***	0.199	0.265	0.438**
	(0.201)	(0.360)	(0.223)	(0.135)	(0.183)	(0.180)
$m^S_{r,t}$	0.298	0.287	0.346	0.413	0.440*	0.509*
	(0.326)	(0.299)	(0.346)	(0.284)	(0.255)	(0.291)
$log(Pop_{r,t})$	-0.134	-0.135*	-0.128	-0.143	-0.139	-0.127
	(0.080)	(0.079)	(0.080)	(0.088)	(0.090)	(0.092)
$log(Urb_{r,t})$	0.277*	0.278*	0.269*	0.284*	0.286*	0.291*
	(0.153)	(0.151)	(0.144)	(0.169)	(0.164)	(0.166)
$log(Hum_{r,t})$	0.818***	0.820***	0.808***	0.857***	0.840***	0.797***
	(0.207)	(0.198)	(0.214)	(0.206)	(0.207)	(0.201)
	(0.125)	(0.143)	(0.127)	(0.110)	(0.107)	(0.116)
Constant	7.436***			7.547***		
	(1.310)			(1.349)		
Observations	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
K-P F-$Test$	N.A	30.90	140.6	N.A	27.64	241.0
Stock Yogo		29.18/16.23	16.38/8.96		29.18/16.23	16.38/8.96
Hansen J (p-value)	N.A	0.388	N.A	N.A	0.232	N.A

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the State level. The specification is described in Eq.(3) and includes all fixed effects. OLS-FE results are provided in col. 1 and 4 and 5; IV results are provided in col. 2, 3, 5 and 6 using the augmented shift-share and gravity-like strategies. Results for college-educated migrants are provided in col. 1 to 3; results for the low-skilled are provided in col. 3 to 6. †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^S_{r,t}$), the log of population ($log(Pop_{r,t})$), the log of urbanization ($log(Urb_{r,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{r,t})$).

Chapter 3

Natives' Attitudes and Immigrants' Unemployment Durations

This chapter is joint work with Sekou Keita (University Clermont Auvergne, IPA).[1] It is currently "revise and resubmit" in *Demography*.

3.1 Introduction

Germany is currently confronted with the challenge of integrating sizable inflows of foreign-born populations which include both economic migrants and refugees.[2] The public debate on the reception of immigrants and asylum seekers sparked very divergent reactions within the German population, ranging from warm welcome demonstrations to violent protestations against this historical surge in foreign born population. Such mixed feelings about immigration are not new and the integration of the foreign-born population in the local labor market has been at the heart of concerns for decades. In this context, a question that is often overlooked is how the attitude of natives affects the integration of immigrants.

[1]We thank Simone Bertoli, Jean-Louis Combes, Vianney Dequiedt, Pascale Phélinas, Anne Viallefont, Pedro Vicente, Ekrame Boubtane, Ababacar Gueye and all the participants to the 2017 SOLE Conference (Raleigh, United States) for their helpful comments . We also thank Herbert Brücker for providing us data on immigration and unemployment rates in Germany. The dataset used in this paper was made available to us by the German Institute for Economic Research (DIW), Berlin.

[2]According to the *Statistisches Bundesamt* (German federal statistical office), 1,226,000 people immigrated to Germany in 2013, an increase of 146,000, or 13%, with respect to 2012. Source:https://www.destatis.de/EN/FactsFigures/SocietyState/Population/Migration/Current.html accessed on July 16, 2016.

The fact that immigrants underperform natives in the labor market has been often observed in the literature (Borjas, 2014). Many potential explanations have been proposed, including immigrants' lower ability, firms' difficulties in properly assessing qualifications obtained in a foreign country, lack of language skills, or discrimination. Empirical evidence suggests that immigrants are indeed discriminated in the labor market (Kaas and Manger, 2012). However, immigrants from different countries of origin do not necessarily face the same obstacles on the destination country labor market. There is a gap in the literature when it comes to exploring the heterogeneity in terms of labor market discrimination against immigrants. In particular, a plausible but unexplored hypothesis is that the different levels of trust that natives associate with immigrants depending on their country of origin might capture origin-specific discrimination. Indeed, trust can condition the willingness to engage in economic transactions with immigrants since trust is a prerequisite for contracts in the absence of complete information (Göran and Hägg, 1994).

While trust levels can affect different labor market outcomes such as wages and job quality, the focus of the present analysis is on unemployment spells for the following reasons. First, we observe an over-representation of immigrants in the unemployed population. Indeed, while the unemployment rate stood at 4.5% for the German native population, it reached 9.2% among immigrants in 2014.[3] We also observe a large heterogeneity by origin country. Second, prolonged unemployment spells bear high costs on society because of poorer health of the unemployed, skill depreciation, forgone tax incomes, slower assimilation, etc.

In this chapter we investigate empirically whether trust levels that Germans associate with the citizens of an immigrants' country of origin influence his or her unemployment duration. The rationale can be conveyed by a standard job search model in which discriminated groups of immigrants, captured by lower levels of trust, receive fewer job offers. As a consequence, immigrants who originate from countries which Germans perceive as less trustworthy end up with lower exit rates out of unemployment.[4]

[3]Source: Eurostat: http://ec.europa.eu/eurostat/statistics-explained/index.php/Migrant_integration_statistics_-_employment accessed on July 16, 2016.

[4]Reducing the arrival rate of jobs offer has two opposed effects. On the one hand, the

We carry out the empirical analysis building on an individual-level panel dataset, the German Socio-Economic Panel (GSOEP). Specifically, we use monthly calendar information to construct labor market activity spells over the period 1984-2012. We then model immigrant's unemployment duration using Cox and Weibull proportional hazard models. We test whether the level of trust that Germans associate with an immigrants' country of origin is a significant determinant of unemployment duration. Our measure for the level of trust is the share of Germans declaring in Eurobarometer surveys that citizens of the country in question are trustworthy. We exploit variation of the levels of trust towards different origin countries both at the national and the regional level. Working at the regional level allows us to control for origin-specific factors which account for the adverse consequences of selection into migration. Indeed, a major drawback of analyzing a self-selected stock-sample is that it excludes potential migrants for whom discrimination is most costly. This in turn can confound the identification of the effect of discrimination on immigrant's labor market outcome and induces a downward bias in the estimated coefficients. In line with this concern, our analysis highlights the importance to overcome the identification challenge posed by varying self-selection patterns across origin countries.

The results of our analysis suggest that natives' attitudes strongly influence the labor market outcomes of immigrants. In particular, our findings indicate that if Germans had the same positive attitudes towards Turkish citizens that they have towards Austrian citizens, Turkish migrants would see their average unemployment duration reduced by three months on average. Our results are robust to alternative specifications and to several definitions of unemployment and levels of aggregation of the variable capturing levels of trust. Furthermore, the results are at odds with a number of alternative explanations.

This chapter is related to several strands of the literature. First, it contributes to the literature investigating the determinants of immigrants' performance on the destination country labor market. The seminal work by Chiswick

unemployment duration of discriminated workers decreases because they become less choosy and reduce their reservation wage. On the other hand, the lower expected number of occasions of leaving unemployment increases immigrant's unemployment duration. van der Berg (1994) shows that under relatively weak conditions the latter effect dominates.

(1978), dealing with the effects of Americanization on the earnings of immigrants, has given rise to a vast literature trying to understand the labor market performance of immigrants in the destination country. Empirical results generally suggest that the ability to engage in social interactions with natives increases the labor market performance of immigrants. For instance, Aldashev et al. (2009) find that language proficiency significantly affects labor market participation, employment probability, and occupational choice of foreigners in Germany. The literature review by Constant and Zimmermann (2009) also suggests that ethnic identities and attitudes seem to have implications for the economic performance of immigrants in the host country. With respect to immigrants' ties to host and origin country, Constant and Zimmermann (2009) find that, conditional upon entering unemployment, those not attached to the host country but rather strongly attached to their country of origin reintegrate more slowly into the German labor market. The authors argue that this group of migrants exerts a relatively low search effort and that it has reservation wages above the level that would imply employment probabilities observed for other groups of migrants. Using the European Social Survey, Bisin et al. (2011) also find that non-EU immigrants in Europe with a strong ethnic identity experience a penalty with respect to labor market outcomes. However, Casey and Dustmann (2010) argue that home country identity and host country identity *per se* are not strong determinants of immigrants' labor market outcomes in Germany. This suggests that the attitude of natives might also play an important role in the economic assimilation of immigrants.

A second strand of the literature related to this chapter investigates how negative attitudes and discrimination affect individuals in their job search process. The economic literature on discrimination distinguishes two forms of discrimination. The first type of discrimination, well-known as taste discrimination, was first described in the seminal paper of Becker (1957). Taste discrimination occurs when members of a minority group are unequally treated regarding members of a majority group despite identical characteristics. The second form of discrimination, namely statistical discrimination, tries to rationalize this unequal treatment. Phelps (1972), Arrow (1973) or Akerlof (1976) analyze discrimination across the informational spectrum and signal theory.

They argue that discrimination arises from the fact that employers lack information about the productivity of job applicants. It becomes therefore rational for them to use observable characteristics like gender or race to infer their productivity, by using the average productivity of the group they belong to. Concerning immigrants, Baker and Benjamin (1994) document the existence of persistent differences in labor market outcomes and small rates of assimilation for different cohorts of immigrants in Canada. The authors argue that the role of discrimination cannot be ruled out. Evidence on important wage differences between natives and immigrants is also found in Germany (Gundel and Peters, 2007; Zibrowius, 2012). The literature suggests that observed wage gaps are not exclusively due to differences in productivity. Indeed, field experiments provide causal evidence that subjective perceptions of recruiters based on race or origin also play an important role (Bertrand and Mullainathan, 2004; Kaas and Manger, 2012; Biavaschi et al., 2013). In the same way, Charles and Guryan (2008) demonstrate that one-quarter of the US black-white wage gap can be directly attributed to discrimination.

Focusing more specifically on natives' self-reported sentiments towards immigrants, Larsen and Waisman (2007) find that in Swedish municipalities with strong negative attitudes, well educated immigrants from developing countries suffer from a sizable income penalty. Furthermore, quasi-experimental evidence indicates that changes in natives attitudes reduce the attractiveness of the host country for immigrants. Based on a household survey, Friebel et al. (2013) find that xenophobic attacks in South Africa against immigrants from neighboring countries decreased the migration intention of household heads in Mozambique. In a similar vein, following a widely documented crime, De Coulon et al. (2016) also identify a significant effect of anti-immigrant attitudes on the intended duration of stay of Romanian migrants in Italy. Gould and Klor (2015) show that the 9/11 terrorist attacks induced a backlash against the Muslim community in the US, slowing their assimilation rate. However, Åslund and Rooth (2005) and Braakmann (2007) find that the variation in attitudes following the attacks of 9/11 did not cause a severe decline in job prospects of Muslims in the Swedish and German labor markets, respectively. In a more general review of the literature on the labor market outcomes of minority groups, Lang and Lehmann

(2012) underline that assessing differences in terms of unemployment duration due to discrimination is rather puzzling from a theoretical perspective. Indeed, since standard job search models ignore the reaction of firms, one might expect that the labor market adjusts through wages and that differences in the exit rate out of unemployment between groups of workers cancel at equilibrium. However, more sophisticated equilibrium search-matching models show that longer unemployment duration for discriminated minority groups can exist at equilibrium. For instance, Rosén (1997) shows that groups of workers with a lower probability of being hired for a job for which they apply end up with lower wages and higher unemployment rate and that this equilibrium is stable. Furthermore, Lang and Lehmann (2012) show that a simple extension of the model by Rosén (1997) can also explain longer unemployment duration and higher turnover for discriminated groups.

A third strand of the literature related to this chapter suggests that perceptions such as trust, rooted in group-specific cultural norms, beliefs, and values, strongly influence economic outcomes. For instance Nunn and Wantchekon (2011) demonstrate that the consequences of the slave trade in terms of mistrust between groups of population in Africa are still observable nowadays. By conditioning the level of social capital and cooperation among agents, inherited trust facilitates economic transactions which is mirrored in macroeconomic outcomes such as economic development (Knack and Keefer, 1997; Fafchamps, 2006) and economic growth (Algan and Cahuc, 2010). Finally Guiso et al. (2009) show that bilateral trust between European countries influences bilateral trade flows, portfolio investments, and direct investments.

The remainder of the chapter is organized as follows. Section 4.2 introduces the data used in the analysis and Section 3.3 presents the descriptive statistics. Section 3.4 presents the empirical strategy and Sections 4.4 and 3.6 the main results and some robustness analysis, respectively. Finally, Section 3.7 draws the main conclusions.

3.2 Data

We use the German Socio-Economic Panel (GSOEP hereafter), the most extensive (still ongoing) tracking survey of private households and persons in Germany. Started in 1984 in the Federal Republic of Germany and extended to the German Democratic Republic in 1990,[5] the GSOEP consists of several samples introduced in different years and targeting specific subgroups of the German population (see appendix Table B8). Samples B and D are the most relevant in the context of our analysis since they oversample households with immigration background.[6]

We combine annual longitudinal biographical information on immigrants with monthly calendar data in order to precisely reconstruct individual labor market activity over time. This information is then matched with German's levels of trust towards the different countries of origin of immigrants in the sample. Information on natives' attitudes is taken from two particular surveys i.e. the Eurobarometer (for the years 1976-1997) and the European Election surveys (for the year 2004).

3.2.1 Labor market activity: The German Socio-Economic Panel

In each wave of the GSOEP, respondents are asked to provide information on their monthly activities of the previous year. Specifically, respondents are asked to choose among 11 different categories the ones corresponding to their main activities in each month. Figure B4 in the appendix provides an example taken from the questionnaire.

We build on this information to generate spells of activity for the year preceding the survey. In particular, we obtain individual-specific monthly activity spells by recoding the 11 activity categories into three labor market statuses: employed, unemployed, and out of labor force. Individuals are classified as employed in months in which they declared being either full-time employed or

[5] With the exception of Berlin, our sample contains only regions from western Germany. This is due to the fact that these regions concentrate the bulk of surveyed immigrants.

[6] Notice that households with immigration background are also present in other samples.

part-time employed. The definition of unemployment is less straightforward because of the difficulty to identify discouraged workers i.e. workers who are not officially registered as unemployed but who are still available for work.[7] It is important to take this particular group into account in our analysis since those individuals could have left the active population precisely because of discrimination. Our sample of analysis consists of individuals officially registered as unemployed as well as individuals who are not officially registered as unemployed but who declare that they are actively looking for work and/or are available for work in the two weeks following the interview. Both pieces of information are taken from the annual biographical questionnaires. We allocate these annual answers to all months of the survey year. Hence, individuals who do not correspond to either activity status are considered out of the labor force and are not included in the analysis. In order to assess the sensitivity of our results to the definition of unemployment, we construct alternative samples with different definitions of unemployment (see Section 3.6.2).

3.2.2 Eurobarometer and European Election surveys

The trust data are taken from different surveys sponsored by the European Commission and designed to measure public opinions on various topics. They were conducted on a representative sample of the total population of age sixteen and older (about 1,000 individuals per country and per year). Specifically, we use waves of Eurobarometer surveys between 1976 and 1997 that collected self-reported trust information of Germans with respect to citizens from 33 countries; we complement this information with the European Election Survey

[7] The International Labour Organization (ILO) resolutions concerning economically active population, employment, unemployment and underemployment adopted by the 13th International Conference of Labour Statisticians, October 1982, paragraph 10, state the following definition: *The unemployed comprise all persons above a specified age who during the reference period were: (i) without work, that is, were not in paid employment or self-employment during the reference period; (ii) currently available for work, that is, were available for paid employment or self-employment during the reference period; and (iii) seeking work, that is, had taken specific steps in a specified recent period to seek paid employment or self-employment.*

which collected similar information in 2004.[8][9] In the Eurobarometer surveys, respondents are asked the following question: *"I would like to ask you a question about how much trust you have in people from various countries. For each, please tell me whether you have a lot of trust, some trust, not very much trust, or no trust at all"*. To construct a measure of bilateral trust from Germans towards other nations, we use the share of positive answers among the total answers i.e. the share of Germans who answered *"very trustworthy"* or *"fairly trustworthy"*. In the waves 1995 and 1997 of the Eurobarometer surveys and in the European Election Survey in 2004, the question is slightly different. The wording of the question is *"do you trust citizens from country X?"*. Only two answers were possible: *"I trust them"* or *"I do not trust them"*. For these years, the share of those who answered *"I trust them"* is used as an indicator of positive opinions.

We are interested in the general level of trust that Germans have towards individuals from different countries of origin. There may be some ambiguity in the interpretation of this measure of trust. Guiso et al. (2009) argue that the correlation with other questions in separate surveys suggests that the level of trust captured in Eurobarometer surveys reflects the subjective probability that a random person from a given country is trustworthy rather than the respondent's ability to identify trustworthy people in a different country.[10]

[8]Unfortunately, Eurobarometer surveys and the European Election Survey do not provide information on the nativity status of respondents. The surveys are designed to capture the opinion of the resident population. Hence it is possible that the opinion of some immigrants were taken into account. According to the 2011 Census, foreigners represented around 10 percent of the German population and diasporas for single origin countries (including German citizens with foreign background i.e. not only migrants strictly speaking) do not reach 4 percent of the total population. Therefore the share of positive opinions we use should not be influenced in a major way by the opinions of migrants.

[9]We provide in the Appendix Table B16 the evidence that our results remain unchanged when focusing only on the 1984 to 1997 period.

[10]Specifically, Guiso et al. (2009) mention a sample of 1,990 individuals who were asked the two following questions: (i) "Suppose that a random person you do not know personally receives by mistake a sum of 1,000 euros that belong to you. He or she is aware that the money belongs to you and knows your name and address. He or she can keep the money without incurring any punishment. According to you what is the probability (a number between zero and 100) that he or she returns the money?" and (ii) "How good are you (very good, good, not very good, not good at all) in detecting people who are trustworthy?" They find that the first question is highly statistically correlated with the measure of trust used in this chapter, but the second one is not.

The factors that shape the perception of the trustworthiness towards citizens from a given country have in common that they are rather stable over time. This is illustrated in Figure B5 in the Appendix. The upper figure shows that the evolution of levels of trust over time is driven by common shocks that do not affect the ranking between countries very much. This appears even more clearly in the lower figure where we partial out year fixed effects that capture shocks such as economic or political conditions in Germany affecting general levels of trust Germans have towards others. This is in line with the literature review on public attitudes towards immigration by Hainmueller and Hopkins (2014) who find that there is little accumulated evidence that immigration-related attitudes are based on personal economic situations and that attitudes seem to be rather driven by symbolic and cultural concerns. Differences in levels of trust are indeed determined by many factors including historical events such as wars, cultural differences, differences in political systems, and the quality of law and its enforcement (Guiso et al., 2009).[11] Hence we obtain our variable of interest by calculating time-invariant origin-specific mean values of the share of Germans who declare that they trust citizens of the origin country in question. We calculate this variable both at the national and regional level. One could be concerned that the variable $Trust$ computed at the regional level reflects statistical noise due to the small number of annual respondents when the Eurobarometer surveys are split between 11 German regions. This concern is mitigated by the fact that the mean value is computed over several waves of Eurobarometer surveys. On average, the mean value of the $Trust$ variable for each region is computed over a sample of 422 individuals.

3.3 Descriptive statistics

Our main sample of analysis is restricted to unemployment spells that do not exceed 48 months.[12] This leaves us with a sample of 108,991 individual-month

[11] Table B14 in Appendix shows the Pearson correlations between our variable $Trust$ and some measures of distances computed by Spolaore and Wacziarg (2009). We clearly see that $Trust$ is highly correlated with indexes of cultural, genetic and religious distance.

[12] We suspect unemployment spells above this threshold to be unusual or potentially artificial (due to early retirement for example). Nonetheless, such observations correspond to less

Table 3.1: Origin countries of immigrants

Country of origin	Obs.	Perc.	Cum.	Survival time 25%	50%	75%	Trust
Turkey	44456	40.79	40.79	7	9	12	0.320
Italy	17101	15.69	56.48	8	12	14	0.558
Poland	13320	12.22	68.64	5	8	10	0.268
Russia	10208	9.37	78.07	7	10	13	0.347
Greece	9421	8.64	86.71	9	13	15	0.596
Spain	3890	3.57	90.28	6	8	12	0.662
Romania	3882	3.56	93.74	6	8	12	0.244
United States	1280	1.17	95.96	3	4	6	0.685
France	1174	1.08	97.83	2	2	3	0.689
Austria	1033	0.95	94.79	4	7	9	0.800
Czech Republic	822	0.79	96.75	13	13	14	0.392
Hungary	796	0.73	98.56	22	22	24	0.517
Netherlands	744	0.68	99.24	13	14	16	0.711
United Kingdom	497	0.46	100.00	9	10	12	0.565
Portugal	324	0.30	99.54	13	15	17	0.597
Total	108991	100.00	Mean	6	9	12	0.406

Source: Author's elaboration on GSOEP panel data, Eurobarometer and European Election Survey data. Survival time is the time elapsed before failure i.e. exit out of unemployment. It is interpreted as the number of months necessary for x% of the unemployed population to find a job. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin.

observations between January 1984 and December 2012. The sample contains 1,580 individuals originating from 15 countries and located in 11 regions (Länder) in Germany. On average, 40.6 percent of Germans declare that they trust citizens that originate from the countries included in our sample of analysis (see Table 3.2). This mean value hides a lot of variability between countries of origin: Table 3.1 shows that only 24.4 percent of Germans perceive Romanians as trustworthy while as much as 80 percent perceive Austrians as trustworthy.

Figure B6 in the appendix reveals that trust towards citizens of a given country also varies greatly between regions. It is interesting to note that the variability across regions is in line with the average at the country level: the

than one percent of the total observations and all results are robust with estimates including unemployment spells above this threshold.

Table 3.2: Descriptives statistics

Variable	Mean	Std.D.	Min	Max	Log-rank Test
Trust	0.406	0.141	0.244	0.800	
Trust(region)	0.396	0.148	0.020	0.813	
Age	1.850	1.005	0	3	5394.600***
Female	0.464	0.499	0	1	924.770***
Nb. Children	1.158	0.165	0	4	954.130***
Education	0.787	0.690	0	2	1784.220***
Married	0.775	0.417	0	1	160.890***
Years since migration	3.292	1.763	0	6	1944.420***
Assistance	0.130	0.336	0	1	18497.350***

Source: Author's elaboration on GSOEP panel data, Eurobarometer and European Election Survey data. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise. The log-rank test compares estimates of the hazard functions of the several groups at each time for categorical variables. The null hypothesis assumes no difference between the survival curves of the different groups.

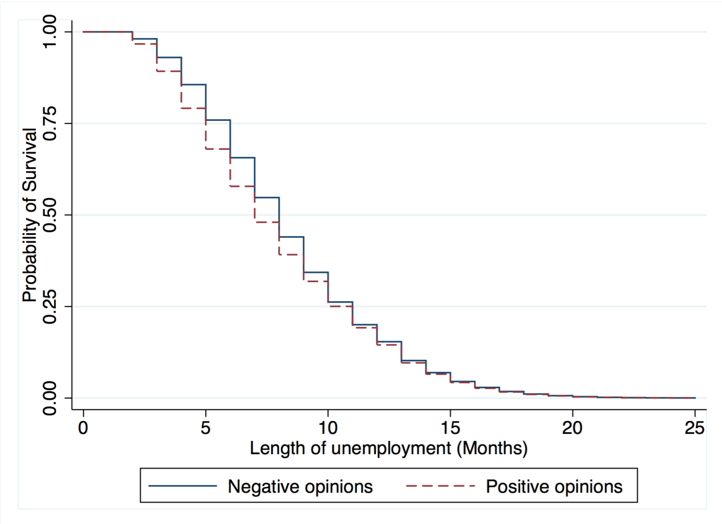
minimum value and maximum of average levels of trust at the regional level are generally within a 20 percentage points range of the country level mean. Hence, the relative level of trust towards citizens of a given country is reflected at the regional level with varying intensity. This can be explained by the heterogeneity within German regions with respect to openness, inherited cultural values and beliefs, which in turn translates into heterogeneity in terms of norms such as family values and attitudes more general (Bertram and Nauck, 1995; Silbereisen and von Eye, 1999; Bertram, 2013; Bertram et al., 2013, for references).

The question at the heart of this chapter is whether varying levels of trust that Germans associate with different countries of origin influence the unemployment duration of immigrants. On average, unemployment spells last for 25.76 months in our main sample of analysis.[13] However, immigrants from different countries of origin experience very different average lengths of unemployment spells. For instance, individuals who originate from Turkey have unemployment spells of 26.53 months on average while the corresponding figure stands at 15.49 months for individuals originating from the US. As can be observed in Table 3.1, longer average spells of unemployment tend to coincide with lower perceived trustworthiness by Germans. Indeed, this table shows that it takes on average 13 months for 75 percent of unemployed immigrants from Russia (low level of trust) to escape unemployment while the corresponding figure for Austrian immigrants (high level of trust) is only nine months.

A similar picture is conveyed by Figure 3.1 which presents survival functions associated with different levels of the variable $Trust$. Indeed, when the share of Germans who perceive the country of origin as trustworthy exceeds 50 percent, the probability of exiting unemployment is higher than when this figure is below 50 percent. However, Table 3.1 also shows that the relationship between the survival time and the level of trust is not systematic. This is not surprising since individuals from different countries of origin are very different in terms of observable characteristics, such as age and years of education, that may influence the duration of unemployment spells (see Table B10 in the appendix). This calls for a more sophisticated analysis that can account for such confounding factors.

[13]This figure is reduced to an average of 21.73 months if we consider only unemployment spells that end with a return to employment during the period of analysis.

Figure 3.1: Kaplan-Meier estimates of immigrant's unemployment duration by natives' attitudes levels



Source: Author's elaboration on GSOEP panel data, Eurobarometers and European Election Survey data. The Kaplan-Meier estimator is a non parametric estimate of the survivor function, which is the probability of failing after the month m (Cleves *et al.*, 2011). The survival function (defined as in Eq. 3.2) reports the probability of remaining unemployed beyond the month m (There was no failure before m). At any month m, it gives the percentage of the population remaining unemployed. Positive and negative opinions refer respectively to the mean of the share of Germans who expressed that they trust or distrust citizens from a given country.

3.4 Empirical Analysis

This section introduces our empirical analysis. After presenting the duration model, we discuss the implications of self-selection of migrants for our analysis, which is due to the stock-sample nature of our dataset. We also discuss how we address this issue by looking at the effect of natives' attitudes at the regional level.

3.4.1 Duration model

We model immigrants' unemployment duration in Germany using survival analysis methods where unemployment duration is reported in months. We define a failure as the transition from unemployment to part-time or full-time employment. Unemployment spells that are not stopped by hiring are treated as right-censored. Let M be a non-negative random variable, called the survival time, denoting the time spent unemployed expressed in number of months. The cumulative distribution of M, $F(m)$ is such as:

$$\begin{aligned} F(m) &= Pr[M \leq m] \\ &= \int_0^m f(s)ds \end{aligned} \tag{3.1}$$

with $f(m)$ its density function. We define $S(m)$, the survival function, as the probability for an individual surviving (remaining unemployed) beyond month m. It is therefore the probability that the individual does not find a job prior to m. The survival function can be written as follows:

$$\begin{aligned} S(m) &= Pr[M \geq m] \\ &= 1 - F(m) \end{aligned} \tag{3.2}$$

This function shows therefore which proportion of unemployed immigrants remains unemployed (i.e. experiences no failure) prior to a given month m. At the heart of duration models, hazard functions focus on the instantaneous probability that an individual finds a job in a given month m, conditional on the fact that he had remained unemployed until this month.[14] Our hazard function is therefore defined such as:

$$h(m) = \lim_{\Delta m \to 0} \frac{Pr(m + \Delta m > M > m | M > m)}{\Delta m} = \frac{f(m)}{S(m)} \tag{3.3}$$

Proportional hazard models (PH) assume that the hazard faced by an individual i, which varies in response to individual's characteristics, is multiplicatively proportional to a baseline hazard $h_0(m)$, faced by all individuals. If we define $\boldsymbol{x}_{im}$ as a vector of covariates with the subscripts i for individual and m for month, our main specification can be written such as:

$$h_i(m|Trust_o, \boldsymbol{x}_{im}) = h_0(m) \exp\{\beta_0 + \beta_1 Trust_o + \boldsymbol{\beta'_x} \boldsymbol{x}_{im}\} \tag{3.4}$$

[14]Months vary from 1 (January 1984) to 348 (December 2012). Individuals enter the analysis in the first month of unemployment.

where $Trust_o$ represents, for an individual i, Germans' trust level towards his or her origin country o. The coefficient of interest β_1 captures the effect of native's attitudes on immigrants' unemployment duration. A positive β_1 means that a higher level of Germans' trust towards the origin country of a given individual increases his instantaneous probability to leave unemployment i.e reduces his expected length of unemployment.[15] Note that our baseline specification takes into account year, seasonal and regional fixed-effects and that $\boldsymbol{x}_{im}$, the full vector of individual characteristics includes age, sex, education, marital status, number of years since migration and whether the unemployed received social assistance.[16] Our empirical analysis uses two different specifications in order to estimate these hazards.[17] These two specifications differ in the assumptions made about the hazard shape over time. First, the semi-parametric Cox-proportional hazard model makes no assumption on time dependency.[18] Thus, hazards can be either an increasing, a decreasing, or a constant function of time. Second, we use the parametric Weibull model which allows us to rewrite the hazard function such as:[19]

$$\begin{aligned} h_i(m|Trust_o, \boldsymbol{x}_{im}) &= h_0(m) \exp\{\beta_0 + \beta_1 Trust_o + \boldsymbol{\beta}'_x \boldsymbol{x}_{im}\} \\ &= pm^{p-1} \exp\{\beta_0 + \beta_1 Trust_o + \boldsymbol{\beta}'_x \boldsymbol{x}_{im}\} \end{aligned} \tag{3.5}$$

with p a parameter, estimated from the data and which models the time dependency of the hazard. If $p > 1$ ($p < 1$) then the hazard is an increasing (decreasing) function of time. Also, our model allows for multiple unemployment spells. In order to avoid time-dependency for the unemployment spells of the same individual over time, we correct the covariance matrix of the esti-

[15]Our tables does not report exponentiated coefficients (hazard ratios) but coefficients. Different from hazard ratios, coefficients have not to be compared with one but with zero.

[16]All these categorical variables have a p-value below 0.01 for the log-rank test of equality. The null hypothesis of the log-rank test assumes equality of survival distributions for each level of a categorical variable. Non parametric estimates of covariates with the Kaplan-Meier estimator are available in appendix Figure B9. All results remain unchanged when the variable assistance is removed from the regression.

[17]Irrespective of the model we use, hazard functions are estimated using a maximum likelihood approach.

[18]Semi-parametric models imply however that the effect of covariates is assumed to take a certain form, by opposition to non parametric models, as the Kaplan-Meier estimator.

[19]The Weibull model is retained against the gamma, log-logistic, log-normal and exponential models, regarding its lower AIC and BIC criteria.

mators by clustering the errors at the individual level (Lin and Wei, 1989).[20]

3.4.2 Empirical strategy

Equation 3.5 is useful for looking at cross-country differences. However, it does not account for origin-specific factors that might influence the exit rate of immigrants out of unemployment. Indeed, an important concern arises from the fact that our analysis builds on a stock-sample of migrants who have chosen to migrate to Germany despite the potential discrimination they would face. Specifically, discrimination might influence the composition of the self-selected group of observed immigrants because different labor market opportunities may lead immigrants from highly discriminated origin countries to be drawn from a different part of the population than their counterparts from less discriminated origin countries.

To the extent that discrimination influences the distribution of wage offers faced by potential migrants, the standard Roy model (Roy, 1951), applied to the analysis of the migration decision by Borjas (1991), predicts that, holding other determinants of individual earnings constant, immigrants originating from a highly discriminated origin country will have on average lower reservation wages compared to immigrants originating from less-discriminated origin countries. This in turn leads to higher acceptance rates of job offers and lower unemployment durations and then, to a downward bias in the estimated coefficient for the variable $Trust_o$.[21] The self-selection of migrants regarding discrimination levels at destination implies that it is crucial to control for origin-specific effects.[22]

[20]Successive failures are assumed to be unordered and of the same type. 40% of the individuals in the baseline sample experienced only one unemployment spell.

[21]Notice that a different argument could lead to the same prediction. For instance, if migrants originating from countries which Germans associate with lower levels of trust also have higher monetary migration costs, the migration in itself would deplete their savings available for the job search period, thus reducing their optimal reservation wage. As a result such migrants would have lower reservation wages as well as shorter unemployment durations.

[22]Lower levels of trust can also influence the selection patterns of immigrants who decide to leave Germany. Individuals suffering the most from discrimination may have greater incentives to leave Germany for another destination or returning back to their origin country. This would imply that the remaining pool of immigrants in Germany is composed of those who are able to mitigate the effects of discrimination due to lower origin-specific trust. Our

In order to overcome the adverse consequences of selection at the national level, we estimate a second equation which considers natives' attitudes at the regional level. In particular, we compute the variable $Trust_{or}$ for 15 origin countries o in each German region r. Our estimated equation becomes:

$$h_i(m|Trust_{or}, \boldsymbol{x}_{im}) = pm^{p-1}\exp\{\beta_0 + \beta_1 Trust_{or} + \boldsymbol{\beta}'_x \boldsymbol{x}_{im}\} \tag{3.6}$$

This specification exploits the variation between origin-region pairs. It allows us to control for unobserved origin-specific factors such as quality of education and self-selection patterns by including origin fixed-effects interacted with year fixed-effects. We expect therefore an increase in our coefficient of interest β_1. It is worth noting that institutions such as collective bargaining and unions can mitigate the downward bias induced by the selection at the national level since they result in a compression of the wage distribution, lower reservation wages in groups of highly discriminated immigrants are not sufficient to fully offset the unemployment effect of lower job offer rates.[23] Indeed, immigrants whose reservation wage is lower than the minimum wage offered by firms cannot increase their exit probability out of unemployment by taking advantage of job offers between their reservation wage and the minimum wage.

Finally, a legitimate concern would be that a similar self-selection effect might occur at the regional level, inducing in a downward bias in the coefficient of $Trust_{or}$. Indeed, as long as we cannot fully correct for self-selection into German regions, our empirical analysis can only determine a conservative estimate of the effect of discrimination on unemployment spells if the migrants in low-trust regions reduce their reservation wage. Figure B10 in the appendix shows that observable characteristics are not systematically different

data reveal that around 10 percent of our sample corresponds to migrants who left Germany between 1984 and 2012. Surprisingly, we found that the mean of the variable $Trust$ is higher for this group (around 50 percent of positive opinions) compared to the mean value for stayers (around 40 percent of positive opinions). This suggests that return migrants are possibly drawn disproportionately more from origin countries towards which Germans express higher trust levels.

[23]Collective bargaining is very common in Germany and labour unions play an important role in the determination of the wages (Franz and Pfeiffer, 2006). This may lead to a compression of wages at the lower end of the distribution even in non-union firms (Blau and Kahn, 1999). Kahn (2000) documents a positive relationship between collective bargaining coverage or union density and low relative employment for less-skilled workers.

for migrants from a given origin country when we compare regions that express relatively higher levels of trust compared to regions that express relatively low levels of trust towards citizens of this origin country. Hence this Figure suggests that migrants in our sample do not systematically self-select into regions that express higher levels of trust towards citizens of their origin country. Although it is not necessarily informative about self-selection on unobserved characteristics, the Figure reduces the concerns regarding self-selection at the regional level.

3.5 Results

This section presents the results of our empirical analysis starting first with the results obtained with the variable $Trust$ at the national level and second with the variable $Trust(region)$ at the regional level. This allows us to discuss the role played by origin-specific unobserved heterogeneity such as self-selection of migrants along the lines discussed in the previous section. Finally, we discuss some threats to identification that could confound our interpretation of the observed correlation between natives' attitudes and immigrant's unemployment duration.

3.5.1 Trust at the national level

Columns (1) and (2) in Table 3.3 report the effect of Germans' trust levels towards the different countries of origin of migrants at the national level. It is worth noticing that these estimates include a full set of individual controls and several sets of fixed-effects. Focusing on our variable of interest, the first two columns also show that regardless of the estimator we employ, a higher level in natives' trust towards a given origin country is associated with a higher instantaneous exit probability out of unemployment for immigrants originating from this country.[24] This effect is significant at the 5% level in both the Cox and Weibull models. The evidence that lower levels of origin-specific trust

[24] The estimated shape parameter $\ln(\rho)$, in Weibull regressions is significantly positive which means that the probability for immigrants to find a job increases with time in unemployment.

are associated with longer unemployment spell for immigrants suggests that immigrants originating from different countries experience diverse barriers to entry when it comes to integrating into the German labor market. Nevertheless, it is important to bear in mind that the coefficients of interest in columns (1) and (2) do not account for the origin-specific self-selection process which, according to a standard Roy model, leads the most discriminated immigrants to be drawn from the lower part of the reservation wage distribution in their home country. Hence these coefficients are potentially downward biased. Regarding individual level variables, it clearly appears that, being aged, being female, having many children, or being married comparatively to single persons, increases immigrants' unemployment duration.[25] Conversely, we observe that conditional upon being unemployed, highly educated immigrants have a higher instantaneous probability of finding a job comparatively to less educated immigrants. The hazard of exiting unemployment also increases with years since immigration, a standard result in the assimilation literature. This is not the case for immigrants receiving financial assistance from the government who experience longer length of unemployment. This result is standard in the literature since benefits may reduce the income gain associated with a transition from unemployment to employment (Bover et al., 2002; Røed and Zhang, 2003).

3.5.2 Trust at the regional level

From columns (3) to (6) in Table 3.3, we estimate the effect of Trust at the regional level. This has the advantage of increasing the variability and then to improve the precision of the estimated parameters. Indeed, additional variability is obtained from the differences observed between the 15 origin countries across 11 regions.[26] Columns (3) and (4) of Table 3.3 show a positive and significant effect of $Trust(region)$ on the probability to leave unemployment.

[25] All the results in this chapter are robust to stratification by gender and available upon request.

[26] We actually use 112 out of 165 possible origin-region pairs because migrants from some origins are not observed in all regions.

Table 3.3: Natives' attitudes and immigrant's unemployment duration. Semi-parametric and parametric estimates.

	(1) Cox Hazard		(2) Weibull Hazard		(3) Cox Hazard		(4) Weibull Hazard		(5) Weibull Hazard		(6) Weibull Hazard	
	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.
Trust	0.193**	(0.087)	0.180**	(0.088)								
Trust(region)					0.231***	(0.085)	0.220**	(0.086)	1.056**	(0.452)	0.400**	(0.163)
Age (ref<25):												
25-34 years old	-0.032	(0.034)	-0.026	(0.034)	-0.033	(0.034)	-0.027	(0.035)	-0.030	(0.035)	-0.028	(0.035)
35-44 years old	-0.120***	(0.040)	-0.124***	(0.041)	-0.121***	(0.040)	-0.125***	(0.041)	-0.127***	(0.041)	-0.129***	(0.041)
45 and + years old	-0.280***	(0.045)	-0.287***	(0.047)	-0.282***	(0.045)	-0.289***	(0.047)	-0.291***	(0.047)	-0.294***	(0.047)
Female (ref is male)	-0.051	(0.036)	-0.055	(0.036)	-0.050	(0.036)	-0.054	(0.036)	-0.056	(0.036)	-0.059	(0.037)
Nb. Children (ref is no child):												
One child	0.055*	(0.030)	0.055*	(0.030)	0.056*	(0.030)	0.056*	(0.030)	0.055*	(0.030)	0.055*	(0.030)
Two children	0.083***	(0.032)	0.063*	(0.034)	0.086***	(0.032)	0.066*	(0.034)	0.066*	(0.034)	0.067**	(0.034)
Three children	0.095**	(0.039)	0.063	(0.041)	0.096**	(0.039)	0.065	(0.040)	0.064	(0.040)	0.066	(0.041)
Four children and +	0.049	(0.054)	0.022	(0.057)	0.051	(0.054)	0.025	(0.057)	0.024	(0.057)	0.027	(0.057)
Female * One child	-0.142***	(0.047)	-0.147***	(0.048)	-0.144***	(0.047)	-0.149***	(0.048)	-0.147***	(0.048)	-0.145***	(0.048)
Female * Two children	-0.106**	(0.048)	-0.093*	(0.049)	-0.109**	(0.048)	-0.096*	(0.050)	-0.096*	(0.050)	-0.091*	(0.050)
Female * Three children	-0.293***	(0.084)	-0.286***	(0.086)	-0.294***	(0.084)	-0.287***	(0.086)	-0.285***	(0.086)	-0.284***	(0.086)
Female * Four children and +	-0.414***	(0.138)	-0.413***	(0.144)	-0.417***	(0.138)	-0.415***	(0.143)	-0.416***	(0.143)	-0.410***	(0.144)
Education (ref is low ISCED):												
Middle ISCED	0.094***	(0.027)	0.099***	(0.027)	0.096***	(0.027)	0.101***	(0.027)	0.101***	(0.027)	0.192**	(0.085)
High ISCED	0.157***	(0.032)	0.172***	(0.032)	0.158***	(0.032)	0.174***	(0.032)	0.174***	(0.032)	0.325***	(0.097)
Married (ref is single)	-0.052**	(0.025)	-0.061**	(0.026)	-0.052**	(0.025)	-0.061**	(0.026)	-0.061**	(0.026)	-0.064**	(0.026)
Years since migration (ref<5):												
5-9 years	0.199***	(0.066)	0.227***	(0.066)	0.200***	(0.066)	0.228***	(0.066)	0.581***	(0.191)	0.227***	(0.066)
10-24 years	0.200***	(0.070)	0.176**	(0.072)	0.202***	(0.070)	0.177**	(0.072)	0.487**	(0.202)	0.177**	(0.072)
15-19 years	0.144**	(0.072)	0.091	(0.073)	0.146**	(0.072)	0.092	(0.073)	0.436**	(0.204)	0.092	(0.073)
20-24 years	0.096	(0.074)	0.026	(0.077)	0.097	(0.074)	0.026	(0.077)	0.285	(0.208)	0.025	(0.076)
25-29 years	0.059	(0.077)	-0.031	(0.079)	0.059	(0.077)	-0.032	(0.079)	0.241	(0.213)	-0.031	(0.079)
30 and +	0.045	(0.080)	-0.073	(0.082)	0.043	(0.080)	-0.075	(0.082)	0.262	(0.217)	-0.071	(0.082)
Assistance	-1.343***	(0.063)	-1.320***	(0.063)	-1.343***	(0.063)	-1.320***	(0.063)	-1.319***	(0.063)	-1.320***	(0.063)
Interactions:												
5-9 years * Trust(region)									-0.970**	(0.440)		
10-14 years * Trust(region)									-0.847*	(0.467)		
15-19 years * Trust(region)									-0.939**	(0.471)		
20-24 years * Trust(region)									-0.720	(0.476)		
25-29 years * Trust(region)									-0.762	(0.484)		
30 and + * Trust(region)									-0.904*	(0.481)		
Middle ISCED * Trust(region)											-0.222	(0.192)
High ISCED* Trust(region)											-0.367*	(0.209)
Constant			-2.513***	(0.131)			-2.534***	(0.131)	-2.835***	(0.220)	-2.608***	(0.141)
Observations	108991		108991		108991		108991		108991		108991	
Individuals	1580		1580		1580		1580		1580		1580	
Failures	71309		71309		71309		71309		71309		71309	
Seasonal fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
Year fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
Regional fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
$ln(\rho)$			0.305***				0.305***		0.306***		0.306***	

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the individual level. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer and European Election Survey data. $ln(\rho)$ is the estimated shape parameter. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

The effect is statistically significant at the 1 and 5% level respectively.[27] It is also worth noticing that in column (5), years since migration reduce the effect of natives' attitudes on immigrants' hazard ratios since the interaction term between the two variables is statistically significant. In the same way, column (6) suggests that the marginal effect of Germans' attitudes is not conditional on the different levels of education at the five percent level of significance.

As discussed in Section 3.4.2, we are concerned that the estimated coefficient of our variable $Trust(region)$ may reflect country-specific factors such as different incentives to migrate that determine the selection of migrants in the population of the country of origin or the fact that different immigrants have faced different quality levels of education in their origin countries. Table B16, column (1), includes therefore origin fixed-effects interacted with year fixed-effects which absorb the effect of all the time-varying origin-specific characteristics which affect immigrant's unemployment duration, and which do not change between German regions. We observe that, while the coefficient remains significant at the one percent level, its magnitude dramatically increases. This is in line with the theoretical intuition that because discrimination reduces the expected gains from migration, migrants from more discriminated origins self-select into the lower parts of the origin country reservation wage distribution. Not accounting for the origin-specific factors reduces therefore significantly the estimated impact of native's attitudes on immigrant's unemployment duration.

In terms of magnitude, if Germans had the same positive attitudes towards Turkish citizens as they have towards Austrian citizens, Turkish migrants would see their average unemployment duration reduced by three months on average.[28] Thus, the effect of natives' attitudes is not just statistical but also an economically significant effect.

[27] Results with $Trust$ and $Trust(region)$ are robust to estimates excluding Turkish immigrants, the largest group of immigrants in Germany. Results are available upon request.

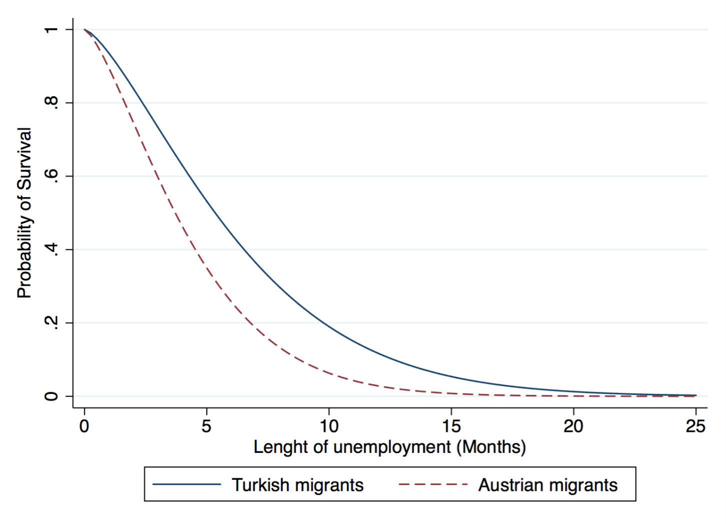
[28] Interpretations in terms of duration are obtained using Weibull estimates in the accelerated failure-time (AFT) metric. AFT model no longer models hazards (as parametric proportional hazards Weibull models) but the logarithm of the duration. Still, the results are perfectly equivalent in the two metrics since $\beta_{AFT} = \frac{-\beta_{PH}}{\rho}$. Results with the AFT metric are not reported here but available upon request.

Table 3.4: Natives' attitudes and immigrant's unemployment duration. Additional fixed-effects and control variables

	(1)		(2)		(3)		(4)		(5)		(6)	
	Weibull Hazard 1984-2012		Weibull Hazard 1984-2012		Weibull Hazard 1991-2012		Weibull Hazard 1984-2012		Weibull Hazard 1991-2012		Weibull Hazard 1991-2012	
	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.
Trust(region)	1.022***	(0.395)	1.142***	(0.413)	1.029***	(0.398)	1.143***	(0.411)	1.004**	(0.397)	1.163***	(0.400)
Immigration rate					1.780	(2.919)						
Stock immigrants (log)							0.019	(0.034)				
Unemployment Rate (monthly)									-0.023**	(0.011)	-0.028***	(0.005)
Age (ref<25):												
25-34 years old	-0.052	(0.036)	-0.042	(0.035)	0.020	(0.049)	-0.042	(0.035)	0.020	(0.049)	0.027	(0.049)
35-44 years old	-0.182***	(0.042)	-0.172***	(0.041)	-0.094*	(0.054)	-0.171***	(0.042)	-0.093*	(0.055)	-0.086	(0.054)
45 and + years old	-0.369***	(0.049)	-0.357***	(0.048)	-0.263***	(0.060)	-0.356***	(0.048)	-0.262***	(0.061)	-0.254***	(0.060)
Female (ref is male)	-0.055	(0.038)	-0.055	(0.037)	-0.034	(0.038)	-0.055	(0.037)	-0.034	(0.038)	-0.035	(0.037)
Nb. Children (ref is no child):												
One child	0.082***	(0.032)	0.088***	(0.031)	0.091***	(0.034)	0.088***	(0.031)	0.090***	(0.034)	0.094***	(0.033)
Two children	0.108***	(0.035)	0.106***	(0.035)	0.118***	(0.036)	0.106***	(0.035)	0.118***	(0.036)	0.116***	(0.037)
Three children	0.127***	(0.042)	0.124***	(0.042)	0.108**	(0.043)	0.125***	(0.042)	0.107**	(0.043)	0.105**	(0.042)
Four children and +	0.072	(0.057)	0.067	(0.054)	0.114**	(0.050)	0.067	(0.054)	0.115**	(0.050)	0.118**	(0.049)
Female * One child	-0.172***	(0.048)	-0.174***	(0.048)	-0.161***	(0.049)	-0.174***	(0.048)	-0.162***	(0.050)	-0.163***	(0.049)
Female * Two children	-0.121**	(0.050)	-0.120**	(0.050)	-0.121**	(0.051)	-0.121**	(0.050)	-0.122**	(0.051)	-0.121**	(0.051)
Female * Three children	-0.313***	(0.086)	-0.317***	(0.086)	-0.307***	(0.096)	-0.318***	(0.086)	-0.306***	(0.096)	-0.300***	(0.096)
Female * Four children and +	-0.428***	(0.141)	-0.410***	(0.138)	-0.485***	(0.147)	-0.409***	(0.138)	-0.481***	(0.146)	-0.454***	(0.143)
Education (ref is low ISCED):												
Middle ISCED	0.073***	(0.027)	0.069**	(0.027)	0.063**	(0.029)	0.069**	(0.027)	0.064**	(0.029)	0.057**	(0.028)
High ISCED	0.122***	(0.034)	0.115***	(0.034)	0.125***	(0.034)	0.115***	(0.034)	0.126***	(0.034)	0.117***	(0.034)
Married (ref is single)	-0.062**	(0.027)	-0.062**	(0.027)	-0.050*	(0.028)	-0.063**	(0.027)	-0.051*	(0.028)	-0.054*	(0.028)
Years since migration (ref<5):												
5-9 years	0.232***	(0.068)	0.189***	(0.065)	0.184**	(0.075)	0.189***	(0.065)	0.182**	(0.075)	0.139*	(0.072)
10-24 years	0.184**	(0.074)	0.142**	(0.071)	0.144*	(0.085)	0.140**	(0.071)	0.143*	(0.084)	0.098	(0.080)
15-19 years	0.099	(0.077)	0.061	(0.073)	0.102	(0.087)	0.057	(0.073)	0.101	(0.087)	0.056	(0.082)
20-24 years	0.049	(0.080)	0.015	(0.077)	0.025	(0.089)	0.011	(0.077)	0.024	(0.089)	-0.016	(0.085)
25-29 years	0.029	(0.082)	-0.013	(0.080)	0.002	(0.090)	-0.016	(0.080)	0.001	(0.090)	-0.044	(0.087)
30 and +	-0.002	(0.084)	-0.033	(0.082)	-0.018	(0.093)	-0.035	(0.082)	-0.019	(0.093)	-0.048	(0.090)
Assistance	-1.298***	(0.062)	-1.310***	(0.062)	-1.304***	(0.062)	-1.310***	(0.062)	-1.301***	(0.062)	-1.317***	(0.062)
Constant	-3.058***	(0.208)	-2.845***	(0.745)	-2.587***	(0.207)	-3.032***	(0.823)	-2.365***	(0.218)	-2.178***	(0.216)
Observations	108991		108991		86596		108991		86596		86596	
Individuals	1580		1580		1301		1580		1301		1301	
Failures	71309		71309		58579		71309		58579		58579	
Seasonal fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
Regional fixed-effects	Yes		No		Yes		No		Yes		No	
Origin fixed-effects	No		Yes		No		Yes		No		Yes	
Origin×Year fixed-effects	Yes		No		Yes		No		Yes		No	
Regional×Year fixed-effects	No		Yes		No		Yes		No		Yes	
$ln(\rho)$	0.337***		0.333***		0.302***		0.333***		0.302***		0.302***	

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the individual level. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer, European Election Survey data, The German Federal Statistical Office and The German Federal Employment Agency. $ln(\rho)$ is the estimated shape parameter. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

Figure 3.2: Predicted survival functions of unemployment



Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer and European Election Survey data. The survival function (defined as in Eq. 3.2) reports the probability of remaining unemployed beyond the month m. At any month, the survival function gives the percent of the population remaining unemployed. These predictions are obtained using the estimated coefficients reported in the column (2) of the Table B16 with the Weibull parametric estimator and using the average value of the control variables.

In order to control for labor market conditions at destination, we introduce in column (2) regional fixed-effects interacted with year fixed-effects which account for all the time-varying region unobserved characteristics that affect immigrant's unemployment duration and which do not vary between origin countries. We prevent therefore our estimations to be biased by yearly heterogeneous dynamics in regional labor markets. Indeed, regions with more favorable labor markets can attract particular groups of immigrants. Still, the coefficient of $Trust(region)$ is highly positive and significant suggesting that negative natives' attitudes towards immigrants hinder their chances of finding a job quicker. Figure 3.2 depicts the predicted survival functions of the immigrant groups with one of the highest and lowest levels of trust namely, Turkish and Austrian immigrants for this last estimate. As expected, the survival function of Austrian immigrants dominates the survival function of Turkish immigrants regardless of the unemployment duration. After seven months of unemployment duration, our model predicts that more than 80 percent of Austrian unemployed would have found a job against less than 65 percent for Turkish immigrants.

3.5.3 Threats to identification

We are concerned that omitted variables could influence simultaneously natives' attitudes towards immigrants and the opportunities for the foreign-born to find a job. First, columns (3) and (4) in Table B16 include therefore annual immigration rates for each migrants' origin countries at the regional level.[29] Indeed, immigration rates can be correlated with natives' attitudes towards a particular origin and also promote (through networks) or deter (through competition) the access to employment for immigrants originating from this country. Our analysis is constrained by the availability of German native population data only after 1991 used to compute immigration rates by origin and region. Thus, we first introduce immigration rates in column (3) with origin fixed-effects interacted with year fixed-effects and restrict our analysis to the 1991-2012 period. The coefficient of $Trust(region)$ remains positive and highly significant. In a second step, in column (4), we introduce the logarithm of the annual stock of immigrants of each origin in each region. This last estimate covers our entire period of analysis. Indeed, regional fixed-effect interacted with time fixed-effects absorb the variation in the size of the natives population over time at the regional level. Still, the effect of natives' attitudes on immigrant's unemployment duration is robust to the introduction of this control variable. Second, from columns (5) to (6), we include monthly unemployment rates at the regional level in order to have a better control of regional market dynamics that can affect both the access of immigrants to local labor market and the native's attitudes. Here again we restrict our analysis to the period 1991-2012. Accounting either for origin fixed-effects interacted with year fixed-effects or for regional fixed-effect interacted with time fixed-effect, the effect of

[29] It is important to note that the definition of immigration rates differs here from what has been used in our analysis so far. Indeed, immigration rates are computed using information on the nationality and not with the origin country of foreign-born. Still, we assume that trends in immigration rates based on nationality reflect the arrival rates of immigrants of different origin country in each region.

$Trust(region)$ on the hazard ratio remains positive and significant at the one percent level. As expected, the coefficient in front of the unemployment rate is negative and significant. It confirms that lower hiring opportunities decrease the likelihood for all individuals to exit unemployment.

Another important source of concern is the fact that variables capturing natives' attitudes might also capture linguistic distance. This entails that our results could reflect the fact that migrants for whom it is more costly to learn German also struggle more to find a job. To tackle this potential issue, we introduce categorical variables capturing self-reported fluency in speaking and writing German language. The results, reported in columns (1) and (2) of appendix Table B15, show that controlling for command of German at the individual level does not modify our main results. One could also be concerned that our results are driven by the fact that migrants from different origins specialize in specific sectors and occupations with varying labor market dynamics. We investigate this possibility by controlling for sector and occupation fixed effects. Results are reported in the appendix Table B15 and suggest that our main results are not due to immigrants clustering in specific occupations and sectors.[30]

Measuring natives' attitudes at the regional level may exacerbate the concern that the ability of a given diaspora to perform in local labor markets could shape native's attitudes at the regional level.[31] This concern is greatly mitigated by the fact that the questions in Eurobarometer surveys ask about trust towards citizen of a given country, not towards migrants in Germany. Nonetheless, it is important to tackle this threat to identification more formally. In order to address the possibility of reverse causality, we rely on an instrumental variable approach, namely the two-stage residual inclusion method (2SRI hereafter), which is widely used to address endogeneity issues in non linear models.[32] Following Terza et al. (2008), our first-stage equation regresses

[30]The changes in coefficient are only due to the fact that the sample size is reduced because of missing values in the additional regressors.

[31]Notice that cross-country estimates are less affected by this issue since it is very unlikely that individual's abilities to perform in local labor markets influence attitudes towards citizen of a particular country at the national level.

[32]The 2SRI estimator corrects for the inconsistency of the estimated parameters obtained with the two-stage least square method (2SLS) applied to non-linear models. While the 2SLS

$Trust(region)$ on its instruments and the control variables at the origin-region level. In a second-stage we include the first-stage residual as an explanatory variable in our equation of interest. We use a proxy for the cultural distance between each German region r and each origin country o as an instrument for $Trust(region)$. As a matter of fact, Guiso et al. (2009) underline that cultural distance is a strong determinant of $Trust$ since individuals tend to have more confidence in people that share their beliefs and values. The exclusion restriction of our IV strategy implies therefore that, conditional on the other covariates included in the regression, cultural distance has no impact on individual's probability to exit unemployment other than through the discrimination channel.

In order to obtain bilateral cultural distances we rely on the World Value Surveys (WVS) which explore values and human beliefs through individual questionnaires conducted in almost 100 countries over the world. Individuals are asked to express their views on several practices. It is crucial to select practices in the questionnaire which do not violate the exclusion restriction of our identification strategy. Therefore we select individual views on homosexuality, abortion, divorce or suicide as instruments. Here, our identification strategy relies on the assumption that tolerance towards homosexuality, abortion, divorce or suicide are very unlikely to influence the individual hazard rate of finding a job other than through cultural distance with natives. For each of these four dimensions we define S_{ir} and S_{io} as the share of individuals either living in the German region r or in the origin country o declaring that the i^{th} dimension is justifiable.[33] The variability of our instrument comes from the fact that, not only different origin countries have different beliefs and values, but also individuals living in different German regions exhibit differences in terms of attitudes towards homosexuality, divorce, suicide or abortion for instance. Indeed, German regions are very heterogeneous with respect to cultural values which translates into heterogeneity in terms of norms such as family val-

and the 2SRI share the same first stage equation, the latter does not replace the endogenous variable by its predicted value but instead, includes the first-stage residuals as additional regressors (Terza et al., 2008).

[33]Data for Austria, Greece and Portugal are not available in the WVS for this particular question.

ues (Bertram and Nauck, 1995; Silbereisen and von Eye, 1999; Bertram, 2013; Bertram et al., 2013, for references). We exploit these intra-regions discrepancies within Germany in order to obtain a bilateral measure of cultural distance between German regions and immigrants' origin countries. Our first measure of cultural distance is such as:

$$DisA_{or} = \sqrt{\sum_i (S_{ir} - S_{io})^2} \tag{3.7}$$

where $i \in \{Homosexuality; Abortion; Divorce; Sucide\}$ is the vector of views on practices. Our second measure of cultural distance is based on four different indicators which measure the probability that two randomly-drawn individuals, one in a given origin country and one in a given region in Germany, have a different opinion on the i^{th} practices:

$$DisB_{ior} = 1 - \Big([S_{ir} * S_{io}] + [1 - S_{ir}] * [1 - S_{io}] \Big) \tag{3.8}$$

Using a principal component analysis, we extract the component $DisPCA_{or}$ that explains most of the variance of the data (56%) and use it as an instrument for $Trust(region)$. We report the results of these estimates in the Table 3.5. We replicate our main result in column (1) excluding Austrian, Greek and Portuguese immigrants from the analysis. It allows us to compare our estimated coefficients across a similar sample, since cultural distance data are not available for these three countries. The coefficient of $Trust(region)$ is still significant and not statistically different from the one reported in Table B16, column(3). In columns (2) we introduce the residuals of the first-stage estimates regressing $Trust(region)$ on $DisA_{or}$.

We see at the bottom of Table 3.5 that the instrument is significant with the expected sign. An increase in the cultural distance between a given German region and a given origin country decreases the share of German that express positive views toward immigrants originating from this country. Moreover the coefficient of $Trust(region)$ remains positive and significant. We find similar results in column (3) using $DisPCA_{o,r}$ as an instrument for $Trust(region)$. However, it is worth noticing that the first-stage residuals in the two last

Table 3.5: Two-stage residuals inclusion method (2SRI)

	(1)		(2)		(3)	
	Weibull		Weibull (2SRI)		Weibull (2SRI)	
	Hazard		Hazard		Hazard	
	β	Std.D.	β	Std.D.	β	Std.D.
Trust(region)	0.995**	(0.403)	1.447***	(0.541)	2.269**	(1.033)
Age (ref<25):						
25-34 years old	-0.048	(0.037)	-0.049	(0.037)	-0.049	(0.037)
35-44 years old	-0.181***	(0.044)	-0.180***	(0.044)	-0.181***	(0.044)
45 and + years old	-0.353***	(0.050)	-0.352***	(0.050)	-0.351***	(0.050)
Female (ref is male)	-0.072*	(0.039)	-0.073*	(0.039)	-0.078**	(0.040)
Nb. Children (ref is no child):						
One child	0.059*	(0.033)	0.058*	(0.033)	0.055*	(0.033)
Two children	0.089**	(0.037)	0.088**	(0.037)	0.086**	(0.037)
Three children	0.117***	(0.044)	0.116***	(0.044)	0.115***	(0.043)
Four children and +	0.073	(0.056)	0.073	(0.056)	0.072	(0.056)
Female * One child	-0.161***	(0.051)	-0.159***	(0.050)	-0.156***	(0.051)
Female * Two children	-0.094*	(0.053)	-0.094*	(0.052)	-0.089*	(0.053)
Female * Three children	-0.331***	(0.093)	-0.329***	(0.093)	-0.322***	(0.093)
Female * Four children and +	-0.445***	(0.143)	-0.445***	(0.143)	-0.436***	(0.144)
Education (ref is low ISCED):						
Middle ISCED	0.084***	(0.029)	0.081***	(0.029)	0.084***	(0.029)
High ISCED	0.138***	(0.038)	0.134***	(0.038)	0.137***	(0.038)
Married (ref is single)	-0.066**	(0.028)	-0.066**	(0.027)	-0.063**	(0.028)
Years since migration (ref<5):						
5-9 years	0.242***	(0.069)	0.240***	(0.069)	0.241***	(0.069)
10-24 years	0.204***	(0.076)	0.201***	(0.076)	0.202***	(0.076)
15-19 years	0.115	(0.078)	0.111	(0.078)	0.110	(0.078)
20-24 years	0.062	(0.082)	0.056	(0.082)	0.055	(0.081)
25-29 years	0.040	(0.084)	0.033	(0.084)	0.030	(0.084)
30 and +	0.019	(0.087)	0.012	(0.087)	0.004	(0.088)
Assistance	-1.247***	(0.063)	-1.247***	(0.063)	-1.243***	(0.063)
First-stage residuals			-0.474	(0.349)	-1.152	(0.838)
Constant	-3.034***	(0.212)	-3.280***	(0.285)	-3.727***	(0.563)
Observations	98213		98213		98213	
Individuals	1403		1403		1403	
Failures	64297		64297		64297	
Seasonal fixed-effects	Yes		Yes		Yes	
Regional fixed-effects	Yes		Yes		Yes	
Origin×Year fixed-effects	Yes		Yes		Yes	
$ln(\rho)$	0.333***	(0.017)	0.333***	(0.016)	0.332***	(0.017)
First-stage:						
$DisInglehart_{or}$		-0.768***	(0.090)			
$DisPCA_{or}$					-0.037***	(0.014)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the individual level. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer, European Election Survey data, The German Federal Statistical Office and The German Federal Employment Agency. $ln(\rho)$ is the estimated shape parameter. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise. Columns (2) and (3) include residuals from the first-stage equations that regress Trust(region) on the control variables and their own instrument $DisInglehart_{r,o}$ and $DisPCA_{or}$ respectively.

columns, which capture determinants of $Trust(region)$ not captured by our instruments are not significant. This strengthens the case for the argument that the variability in German attitudes captured in our regressions are not shaped by the local labor market performance of immigrants. In other words, unobserved origin-region characteristics correlated with $Trust(region)$, such as the average labor market performance of migrants from a given origin country, are not driving individuals' hazard rates of exiting unemployment.[34] This is in line with our observation that $Trust$ is stable over time because it is determined by historical legacy and deeply rooted cultural differences as discussed in Section 4.2.

3.6 Robustness

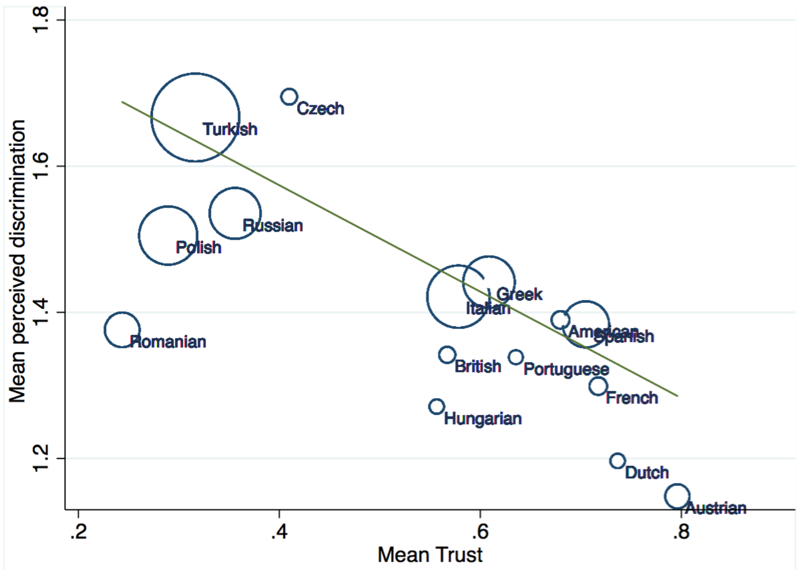
3.6.1 Perceived discrimination

A legitimate question is whether different levels of trust towards citizens of a given country translate to immigrants that originate from this country. Indeed, it could be possible that well integrated immigrants are perceived differently than their fellow countrymen who live in their country of origin. We explore this point by observing the correlation between the level of trust Germans have towards citizens of a given country of origin and the perception of these immigrants with respect to discrimination due to their origin. This analysis at the individual level has also the advantage of exploiting a different source of variability that complements our previous group-level analysis. Specifically, we use a variable that captures the answer to the following question: *"how often have you experienced disadvantages in the last two years because of your origins?"*. The corresponding variable takes the value of 0 if the response is *never*, 1 if the response is *seldom*, and 2 if the response is *often*. Figure 3.3 shows a strong and negative correlation between the variable $Trust$ and the perceived discrimination variable: immigrants who originate from countries that Germans associate with lower levels of trust declare that they are more often

[34]It is important to note that, included one by one in our main specification as explanatory variables, our two instruments are not significant which supports the validity of the exclusion restriction.

discriminated due to their origin.[35] Lower levels of trust of Germans towards a particular group of immigrants seem therefore to be strongly correlated with immigrants' perceived discrimination.

Figure 3.3: Natives' attitudes and immigrants' perceived discrimination



Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer and European Election Survey data. The size of the markers is scaled according to the size of each diaspora in the sample. Mean Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Mean perceived discrimination by origin country refers to the average responses to the question *"how often have you experienced disadvantages in the last two years because of your origins?"*. The variable takes the value of 0 if the response is *never*, 1 if the response is *seldom*, and 2 if the response is *often*. The line plots the linear predictions of the regression of Mean perceived discrimination on Mean Trust.

In this Section we investigate empirically whether higher levels of perceived discrimination are associated with longer unemployment durations among immigrants. We use the information on 1376 immigrants originating from 24 different countries.[36] We also look at how the feeling of belonging to the German nationality influences integration into the local labor market. Indeed,

[35]The correlation coefficient between the two variables stands at -0.17 and is statistically significant at the one percent level.

[36]Descriptive statistics by origin countries and at the global level are available in appendix Tables B11 and B12.

respondents in the GSOEP are asked to reply to the following question: *"How much do you feel like a German?"*. Our German identity variable takes therefore the value of 0 if the response is *completely*, 1 if the response is *mostly*, 2 if the response is *in some respects*, 3 if the response is *barely* and 4 if the response is *not at all*. The idea behind this second set of estimations is to say that if the effect of the variable $Trust$ really captures a discrimination process, then we should find a negative effect of feeling discriminated but no strong effect of feeling German. In other words, the negative impact observed on immigrant's unemployment durations has to come from natives' attitudes and not from the attitude of immigrants. It is worth noting that (time-varying) perceived discrimination and unemployment duration certainly strongly affect each other and that these results should be considered with great caution. However the correlations that we present below are interesting in the sense that they allow to complement the main results and explore alternative explanations.

Table 3.6 presents the effect of perceived discrimination on immigrant's unemployment duration from columns (1) to (3). After controlling for potentially confounding factors through the inclusion of origin fixed effects we see that higher levels of perceived discrimination are negatively associated with the hazard rate i.e the expected length of unemployment for immigrants is higher for migrants who declare having been discriminated due to their origin. Coefficients are fairly stable regardless of the different specifications. Taking into account either time-varying origin-specific characteristics in column (2) or time-varying region-specific characteristics in column (3), we observe that immigrants who often feel discriminated also experience longer unemployment durations comparatively to immigrants who never feel discriminated. These patterns are less straightforward when it comes to German identity in columns (4), (5) and (6). Despite the fact that in the first specification, higher levels of German identity seem to be associated with lower length of unemployment among immigrants in columns (4). This effect is no longer significant in more sophisticated specifications. After controlling for different forms of unobserved heterogeneity, feeling more or less German is not strongly associated with immigrant's unemployment durations. This result is in line with the intuition that, while immigrants can have a strong feeling of belonging to Germany, negative attitudes from natives can still deteriorate their entrance in the local labor-market. Longer unemployment durations for some particular group of immigrants seem therefore to come from native's attitudes more than immigrant's behaviours or feelings. This is in line with the evidence given by (Casey and Dustmann, 2010) that host country identity is not a strong determinant of immigrants' success in the labor market.

3.6.2 Alternative definitions of unemployment

In this section we look at the sensitivity of our results when we change the definition of unemployment. Recall that the baseline sample which has been used so far is the **Sample 3**. Column (3) in Table 3.7 reports therefore the previous findings of the chapter for easier comparison.

First, we restrict our analysis to **Sample 1** which only includes immigrants officially registered as unemployed. We find that higher values of $Trust(region)$ are associated with higher exit rates out of unemployment. In addition, with the **Sample 2**, our main results do not change. Adding individuals looking for work raises the size of the sample but does not modify the size of the coefficient of interest. When we add to **Sample 2** immigrants not actively looking for a job but declaring to be available for work in the two weeks following the interview, the coefficient of interest increases dramatically in **Sample 3**. The intuition is that excluding these individuals from the analysis leads to a downward bias of the estimates since discouraged immigrants workers could have left the labor market due to discrimination in the hiring process. As a matter of fact, these individuals that we identify as discouraged workers are more likely to be subject to discrimination and may have left the labor market for this particular reason. Recoding short periods out of the labor force in the last column, our results stay stable.[37] Indeed, in **Sample 4**, we recode inactivity spells of less than one year between two employment spells as employment. In most occurrences this corresponds to holidays or maternity leaves.[38] We also

[37] A short exit of the active population corresponds to spells of inactivity shorter than one year.

[38] We noticed that most of the short spells of inactivity correspond to months that are typically used for vacation in Germany such as July and August.

Table 3.6: Perceived discrimination, German identity and immigrants' unemployment duration.

	(1) Weibull Hazard		(2) Weibull Hazard		(3) Weibull Hazard		(4) Weibull Hazard		(5) Weibull Hazard		(6) Weibull Hazard	
	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.
Discrimination (ref is never):												
Seldom	-0.051***	(0.017)	-0.045***	(0.017)	-0.044**	(0.017)						
Often	-0.078**	(0.039)	-0.061	(0.039)	-0.062	(0.038)						
German identity (ref is completely):												
Mostly							-0.071	(0.054)	-0.040	(0.053)	-0.048	(0.055)
In some respects							-0.102**	(0.052)	-0.029	(0.053)	-0.031	(0.053)
Barely							-0.121**	(0.056)	-0.019	(0.058)	-0.023	(0.058)
Not at all							-0.145***	(0.055)	-0.042	(0.059)	-0.036	(0.058)
Age (ref<25):												
25-34 years old	0.135**	(0.056)	0.134**	(0.056)	0.135**	(0.056)	-0.089**	(0.042)	-0.117***	(0.045)	-0.135***	(0.044)
35-44 years old	0.054	(0.061)	0.012	(0.061)	0.026	(0.061)	-0.119**	(0.051)	-0.204***	(0.054)	-0.221***	(0.054)
45 and + years	-0.039	(0.065)	-0.093	(0.066)	-0.078	(0.065)	-0.383***	(0.062)	-0.478***	(0.066)	-0.476***	(0.066)
Female (ref is male)	-0.021	(0.033)	-0.018	(0.034)	-0.018	(0.034)	-0.086	(0.053)	-0.112**	(0.055)	-0.105*	(0.055)
Nb. Children (ref is no child):												
One child	0.061**	(0.030)	0.064**	(0.031)	0.069**	(0.031)	0.045	(0.046)	0.057	(0.046)	0.067	(0.046)
Two children	0.074**	(0.033)	0.096***	(0.034)	0.102***	(0.034)	0.013	(0.050)	0.043	(0.050)	0.056	(0.052)
Three children	0.030	(0.037)	0.086**	(0.039)	0.087**	(0.040)	0.000	(0.066)	0.037	(0.066)	0.046	(0.066)
Four children	0.010	(0.048)	0.044	(0.048)	0.055	(0.048)	-0.129	(0.092)	-0.071	(0.096)	-0.071	(0.091)
Female * One child	-0.152***	(0.047)	-0.148***	(0.047)	-0.154***	(0.047)	-0.158**	(0.072)	-0.148**	(0.071)	-0.153**	(0.072)
Female * Two children	-0.169***	(0.052)	-0.196***	(0.053)	-0.196***	(0.053)	-0.139*	(0.078)	-0.141*	(0.079)	-0.145*	(0.080)
Female * Three children	-0.324***	(0.100)	-0.381***	(0.098)	-0.368***	(0.098)	-0.180*	(0.109)	-0.157	(0.109)	-0.172	(0.107)
Female * Four children and +	-0.370**	(0.165)	-0.417***	(0.158)	-0.423***	(0.162)	-0.208	(0.221)	-0.180	(0.221)	-0.200	(0.218)
Education (ref is low ISCED):												
Middle ISCED	0.094***	(0.028)	0.074***	(0.028)	0.068**	(0.028)	0.069*	(0.036)	0.062*	(0.035)	0.053	(0.035)
High ISCED	0.141***	(0.034)	0.115***	(0.034)	0.106***	(0.035)	0.154***	(0.055)	0.124**	(0.057)	0.120**	(0.056)
Married (ref is single)	-0.017	(0.027)	-0.003	(0.027)	-0.003	(0.027)	-0.029	(0.039)	-0.019	(0.040)	-0.020	(0.040)
Years since migration (ref<5):												
5-9 years	0.125*	(0.067)	0.130*	(0.067)	0.079	(0.063)	0.088	(0.081)	0.109	(0.081)	0.101	(0.080)
10-14 years	0.087	(0.073)	0.125	(0.076)	0.055	(0.070)	0.032	(0.084)	0.076	(0.087)	0.054	(0.084)
15-19 years	0.076	(0.076)	0.139*	(0.080)	0.055	(0.073)	-0.098	(0.088)	-0.047	(0.091)	-0.079	(0.088)
20-24 years	0.020	(0.080)	0.093	(0.082)	0.015	(0.077)	-0.103	(0.091)	-0.069	(0.095)	-0.081	(0.092)
25-29 years	-0.054	(0.081)	0.014	(0.084)	-0.049	(0.079)	-0.175*	(0.097)	-0.136	(0.103)	-0.158	(0.101)
30 and +	-0.096	(0.083)	-0.026	(0.086)	-0.089	(0.082)	-0.297***	(0.114)	-0.217*	(0.125)	-0.267**	(0.121)
Assistance	-1.357***	(0.062)	-1.348***	(0.061)	-1.353***	(0.061)	-1.387***	(0.121)	-1.412***	(0.120)	-1.405***	(0.120)
Constant	-2.138***	(0.130)	-2.389***	(0.146)	-2.254***	(0.225)	-2.638***	(0.237)	-3.016***	(0.285)	-2.676***	(0.755)
Observations	75733		75733		75733		41977		41977		41977	
Individuals	1376		1376		1376		1236		1236		1236	
Failures	51053		51053		51053		25421		25421		25421	
Year fixed-effects	Yes		No		No		Yes		No		No	
Regional fixed-effects	Yes		Yes		No		Yes		Yes		No	
Seasonal fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
Origin fixed-effects	No		No		Yes		No		No		Yes	
Origin×Year fixed-effects	No		Yes		No		No		Yes		No	
Regional×Year fixed-effects	No		No		Yes		No		No		Yes	
$ln(\rho)$	0.267***	(0.018)	0.281***	(0.019)	0.282***	(0.019)	0.389***	(0.024)	0.418***	(0.025)	0.417***	(0.025)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors adjusted for clustering at the individual level. Source: Author's elaborations on SOEP panel data over 1984-2012. $ln(\rho)$ is the estimated shape parameter. Discrimination refers to the question *"how often have you experienced disadvantages in the last two years because of your origins?"*. The variable takes the value of 0 if the response is *never*, 1 if the response is *seldom*, and 2 if the response is *often*. German identity refers to the question *"How much do you feel like a German?"*. German identity takes the value of 0 if the response is *completely*, 1 if the response is *mostly*, 2 if the response is *in some respects*, 3 if the response is *barely* and 4 if the response is *not at all*. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

recode short inactivity spells between two spells of unemployment as unemployment and extend unemployment spell duration when the short period of inactivity is framed between an unemployment spell and an employment spell. The rationale is that these individuals may also be discouraged unemployed workers since they took a job as soon as they were offered an opportunity. Finally, we extend unemployment spell duration when the short inactivity spell is preceded by an employment spell and followed by an unemployment spell. We assume that during this short time out of the labor force the individual was looking for a job or at least was available for work. The coefficient is robust and keeps both its magnitude and its statistical significance with this last sample.

Table 3.7: Robustness to alternative definitions of unemployment

	(1)		(2)		(3)		(4)	
	Sample 1		**Sample 2**		**Sample 3**		**Sample 4**	
	Weibull Hazard		Weibull Hazard		Weibull Hazard		Weibull Hazard	
	β	Std.D.	β	Std.D.	β	Std.D.		
Trust(region)	0.665**	(0.317)	0.659**	(0.332)	1.022***	(0.395)	1.022***	(0.392)
Age (ref<25):								
25-34 years old	-0.079**	(0.036)	-0.050	(0.036)	-0.052	(0.036)	-0.054	(0.036)
35-44 years old	-0.219***	(0.044)	-0.194***	(0.044)	-0.182***	(0.042)	-0.184***	(0.042)
45 and + years old	-0.400***	(0.050)	-0.374***	(0.050)	-0.369***	(0.049)	-0.369***	(0.049)
Female (ref is male)	-0.029	(0.036)	-0.044	(0.037)	-0.055	(0.038)	-0.054	(0.037)
Nb. Children (ref is no child):								
One child	0.037	(0.031)	0.053*	(0.031)	0.082***	(0.032)	0.081**	(0.032)
Two children	0.095***	(0.033)	0.103***	(0.034)	0.108***	(0.035)	0.106***	(0.035)
Three children	0.114***	(0.042)	0.116***	(0.042)	0.127***	(0.042)	0.125***	(0.042)
Four children and +	0.082	(0.057)	0.079	(0.059)	0.072	(0.057)	0.071	(0.056)
Female * One child	-0.081*	(0.048)	-0.097**	(0.049)	-0.172***	(0.048)	-0.169***	(0.048)
Female * Two children	-0.060	(0.051)	-0.089*	(0.051)	-0.121**	(0.050)	-0.119**	(0.049)
Female * Three children	-0.275***	(0.103)	-0.363***	(0.111)	-0.313***	(0.086)	-0.299***	(0.086)
Female * Four children and +	-0.229	(0.165)	-0.292*	(0.162)	-0.428***	(0.141)	-0.429***	(0.140)
Education (ref is low ISCED):								
Middle ISCED	0.096***	(0.028)	0.083***	(0.028)	0.073***	(0.027)	0.073***	(0.027)
High ISCED	0.153***	(0.035)	0.141***	(0.035)	0.122***	(0.034)	0.121***	(0.034)
Married (ref is single)	-0.002	(0.028)	-0.015	(0.028)	-0.062**	(0.027)	-0.059**	(0.027)
Years since migration (ref<5):								
5-9 years	0.166**	(0.067)	0.156**	(0.068)	0.232***	(0.068)	0.224***	(0.068)
10-140 years	0.111	(0.072)	0.107	(0.073)	0.184**	(0.074)	0.176**	(0.074)
15-19 years	0.054	(0.075)	0.043	(0.076)	0.099	(0.077)	0.093	(0.077)
20-24 years	0.028	(0.078)	0.023	(0.078)	0.049	(0.080)	0.040	(0.080)
25-29 years	0.001	(0.080)	-0.011	(0.081)	0.029	(0.082)	0.022	(0.082)
30 and +	-0.003	(0.083)	-0.021	(0.083)	-0.002	(0.084)	-0.010	(0.084)
Assistance	-1.341***	(0.062)	-1.332***	(0.063)	-1.298***	(0.062)	-1.298***	(0.062)
Constant	-2.171***	(0.196)	-2.316***	(0.203)	-3.058***	(0.208)	-3.051***	(0.207)
Observations	82774		88840		108991		109765	
Individuals	1243		1333		1580		1580	
Failures	56988		60031		72083		72083	
Seasonal fixed-effects	Yes		Yes		Yes		Yes	
Regional fixed-effects	Yes		Yes		Yes		Yes	
Origin × Year fixed-effects	Yes		Yes		Yes		Yes	
$ln(\rho)$	0.241***	(0.017)	0.277***	(0.017)	0.337***	(0.016)	0.337***	(0.016)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the individual level. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer and European Election Survey data. $ln(\rho)$ is the estimated shape parameter. Each sample corresponds to a different definition of unemployment. Trust(region) is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin and by German regions. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

3.7 Conclusions

In the debate on the integration of immigrants in the labor market of the destination country, the role of the attitudes of natives has often been overlooked. In particular, varying attitudes across German regions towards immigrants from different countries of origins might contribute to explain observed heterogeneity in terms of immigrants' labor market performance.

In this chapter we investigate how natives' attitudes relate to the unemployment duration of immigrants. Our empirical analysis finds that positive German attitudes are associated with shorter unemployment duration for migrants. By combining data from the GSOEP and Eurobarometer surveys, our estimates indicate that if Germans had the same positive attitudes towards Turkish citizens that they have towards Austrian citizens, the Turkish migrants would see their average unemployment duration reduced by three months. This result is particularly important given the large number of migrants originating from developing countries who are expected to come to Germany and other high-income countries in upcoming years. It underlines that the assimilation of foreigners at destination is not only the responsibility of newcomers but also of the native population. We find that this effect is robust to different specifications and alternative definition of unemployment.

Reducing negative attitudes in migrants' host countries towards foreign-born should be at the heart of integration policies since it affects returns to education and the incentives for immigrants to invest in human capital at destination. This aspect has been stressed as crucial in the assimilation process (Borjas, 2014). Integrating foreign-born is particularly important given the direct costs of unemployment for host societies and the opportunity cost of an untapped workforce which could better contribute to the economic growth if it was employed at its full potential. Raising awareness on these issues has lead policy makers to introduce anti-discrimination policies with the major goal to overcome the negative effects of discrimination on immigrants' labor market outcomes. As a matter of fact, Germany ranked 22nd out of 38 in the Migration Integration Policy index which measures the effort of the integration of immigrants made by OECD countries.[39] Also, if natives' attitudes reflect cultural, historical and political differences, then the main focus should be on public beliefs and resentment about immigrants from different origins.

[39]http://www.mipex.eu/anti-discrimination.

References

Akerlof, G. (1976). The economics of caste and of the rat race and other woeful tales. *Quarterly Journal of Economics*, 94(1):599–187.

Aldashev, A., Gernandt, J., and Thomsen, S. L. (2009). Language usage, participation, employment and earnings: Evidence for foreigners in West Germany with multiple sources of selection. *Labour Economics*, 16(3):330–341.

Algan, Y. and Cahuc, P. (2010). Inherited trust and growth. *The American Economic Review*, 100(5):2060–2092.

Arrow, K. (1973). The theory of discrimination in: O. A. Ashenfelter and A. Rees, eds., *Discrimination in labor markets.* Princeton, N.J: Princeton University Press, 3–33.

Åslund, O. and Rooth, D.-O. (2005). Shifts in attitudes and labor market discrimination: Swedish experiences after 9-11. *Journal of Population Economics*, 18(4):603–629.

Baker, M. and Benjamin, D. (1994). The performance of immigrants in the Canadian labor market. *Journal of Labor Economics*, 12(1):369–405.

Becker, G. S. (1957). *The economics of discrimination.* Chicago: the University of Chicago Press.

Bertram, H. (2013). *Das Individuum und seine Familie: Lebensformen, Familienbeziehungen und Lebensereignisse im Erwachsenenalter*, volume 4. Springer-Verlag.

Bertram, H. and Nauck, B. (1995). Kinder in deutschland. *Lebensverhältnisse von Kindern im Regionalvergleich. DJI-Familiensurvey, Bd*, 5.

Bertram, H., Bayer, H., and Bauereiß, R. (2013). *Familien-Atlas: Lebenslagen und Regionen in Deutschland: Karten und Zahlen.* Springer-Verlag.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94(4):991–1013.

Biavaschi, C., Giulietti, C., and Siddique, Z. (2013). The economic payoff of name Americanization. *IZA Discussion Paper No.7725*. Institute for the Study of Labor (IZA), Bonn, Germany.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Ethnic identity and labour market outcomes of immigrants in Europe. *Economic Policy*, 26(65):57–92.

Blau, F. D and Kahn, L. M. (1999). *Institutions and laws in the labor market.* Handbook of Labor Economics, 3A:1399–1461.

Borjas, G. J. (1991). Immigration and self-selection. In *Immigration, trade, and the labor market*, pages 29–76. University of Chicago Press.

Borjas, G. J. (2014). *Immigration economics.* Harvard University Press.

Bover, O., Arellano, M., and Bentolila, S. (2002). Unemployment duration, benefit duration and the business cycle. *The Economic Journal*, 112(479):223–265.

Braakmann, N. (2007). The impact of september 11th, 2001 on the job prospects of foreigners with Arab background: Evidence from German labor market data. University of Lüneburg Working Paper Series in Economics.

Card, D., Dustmann, C., and Preston, I. (2005). Understanding attitudes to immigration: The migration and minority module of the first European social survey. CReAM Discussion Paper No 03/05.

Casey, T. and Dustmann, C. (2010). Immigrants' identity, economic outcomes and the transmission of identity across generations. *The Economic Journal*, 120(542):31–51.

Charles, K., K., and Guryan, J. (2008). Prejudices and wages: An empirical assessment of Becker's The Economics of Discrimination.. *Journal of Political Economy*, 116(5):773–809.

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *The Journal of Political Economy*, 86(5):897–921.

Cleves M., Gutierrez G. G., Gould, W. and Marchenko, Y. V. (2010). An introduction to survival analysis using Stata. Third edition. Stata Press Publication.

Constant, A. F. and Zimmermann, K. F. (2009). Migration, ethnicity and economic integration. *IZA Discussion Paper No.4620*. Institute for the Study of Labor (IZA), Bonn, Germany.

De Coulon, A., Radu, D., and Friedrich Steinhardt, M. (2016). Pane e cioccolata: The impact of native attitudes on return migration. *Review of International Economics*, 24(2):253–281.

Fafchamps, M. (2006). Development and social capital. *The Journal of Development Studies*, 42(7):1180–1198.

Franz, W. and Pfeiffer, F. (2006). Reasons for wage rigidity in Germany. *Labour*, 20(2):255–284.

Friebel, G., Gallego, J. M., and Mendola, M. (2013). Xenophobic attacks, migration intentions, and networks: Evidence from the South of Africa. *Journal of Population Economics*, 26(2):555–591.

Göran, P. and Hägg, T. (1994). The economics of trust, trust-sensitive contracts, and regulation. *International Review of Law and Economics*, 14(4):437–451.

Gould, E. D. and Klor, E. F. (2015). The long-run effect of 9/11: Terrorism, backlash, and the assimilation of Muslim immigrants in the West. *The Economic Journal.*, forthcoming.

Guiso, L., Sapienza, P., and Zingales, L. (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics*, 124(3):1095–1131.

Gundel, S. and Peters, H. (2007). Assimilation and cohort effects for German immigrants. *SOEP paper No.64*.

Hainmueller, J. and Hopkins, D. J. (2014). Public attitudes toward immigration. *Annual Review of Political Science*, 17()1:225–249.

Kaas, L. and Manger, C. (2012). Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review*, 13(1):1–20.

Kahn, L. M. (2000). Wage inequality, collective bargaining, and relative employment from 1985 to 1994: Evidence from fifteen OECD countries. *The Review of Economics and Statistics*, 82(4):564–579.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, 112(4): 1251–1288.

Lang, K. and Lehmann, J-Y. (2012). Racial discrimination in the labor markert: Theory and empirics. *Journal of Economic Literature*, 50(4): 959–1006.

Larsen, B. and Waisman, G. (2007). Do attitudes towards immigrants matter? CBS Working Paper.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for Cox proportional hazards model. *Journal of the American Statistical Association*, 84(1): 1047–1078.

Mayda, A. M. (2006). Who is against immigration? A cross-country investigation of individual attitudes toward immigrants. *The Review of Economics and Statistics*, 88(3):510–530.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *The American Economic Review*, 101(7):3221–3252.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(1):659–661.

Røed, K. and Zhang, T. (2003). Does unemployment compensation affect unemployment duration?. *The Economic Journal*, 113(484):190–206.

Rosén, A. (1997). An equilibrium search-matching model of discrimination. *European Economic Review*, 41(1):1589–1613.

Roy, A., D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146.

Silbereisen, R. K. and von Eye, A. (1999). *Growing up in times of social change*, volume 7. Walter de Gruyter.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Eonomics*, 124(2):469–529.

Terza, J., Bazu, A., and Rathouz, P. (2008). A two stage residual inclusion estimation: Adressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3):531–543.

van den Berg, G. J. (1994). The effects of changes of the job offer arrival rate on the duration of unemployment. *Journal of Labor Economics*, 12(3):478–498.

Zibrowius, M. (2012). Convergence or divergence? Immigrant wage assimilation patterns in Germany. *SOEP paper No.479*.

Appendix to chapter 3

Figure B4: Calendar data from the SOEP

```
+--------------------------------------------------------------------------+
| 74.     And now please think about the entire previous year, in other    |
|         words about 1993:  We have made a sort of calendar.  On the      |
|         left, we have written things that could have happened last year. |
|         Please go through the entire list and check each month, in       |
|         which, for example, you were employed or unemployed, etc.        |
|         Please make sure you answer for each month.                      |
|                                                                          |
| 1993                                          J F M A M J J A S O N D    |
| full-time employment, job creation measure    _ _ _ _ _ _ _ _ _ _ _ _    |
| Short-time work                               _ _ _ _ _ _ _ _ _ _ _ _    |
| part-time or occasionally employed            _ _ _ _ _ _ _ _ _ _ _ _    |
| vocational training, education, retraining    _ _ _ _ _ _ _ _ _ _ _ _    |
| registered unemployed                         _ _ _ _ _ _ _ _ _ _ _ _    |
| retired, early retirement                     _ _ _ _ _ _ _ _ _ _ _ _    |
| maternity leave                               _ _ _ _ _ _ _ _ _ _ _ _    |
| in school/college                             _ _ _ _ _ _ _ _ _ _ _ _    |
| military/civilian service                     _ _ _ _ _ _ _ _ _ _ _ _    |
| housewife/househusband                        _ _ _ _ _ _ _ _ _ _ _ _    |
| other, namely  __________                     _ _ _ _ _ _ _ _ _ _ _ _    |
+--------------------------------------------------------------------------+
```

Source: Desktop Companion to the German Socio-Economic Panel (GSOEP).

Table B8: GSOEP samples

Sample A	German residents of West Germany, stared in 1984.
Sample B	Foreigners in West Germany with Turkish, Greek, Yugoslavian, Spanish or Italian household head. Sample B is oversampled and started with in 1984.
Sample C	Est German households, started in June 1990.
Sample D	Immigrants sample, started in 1995.
Sample E	Refreshment sample, started in 1998. A new sample was selected from the population of private households in Germany.
Sample F	Innovation sample, new households added in 2000.
Sample G	High Income Sub-sample, households with a monthly income of at least DM 7,500 (EURO 3,835). Started in 2002.

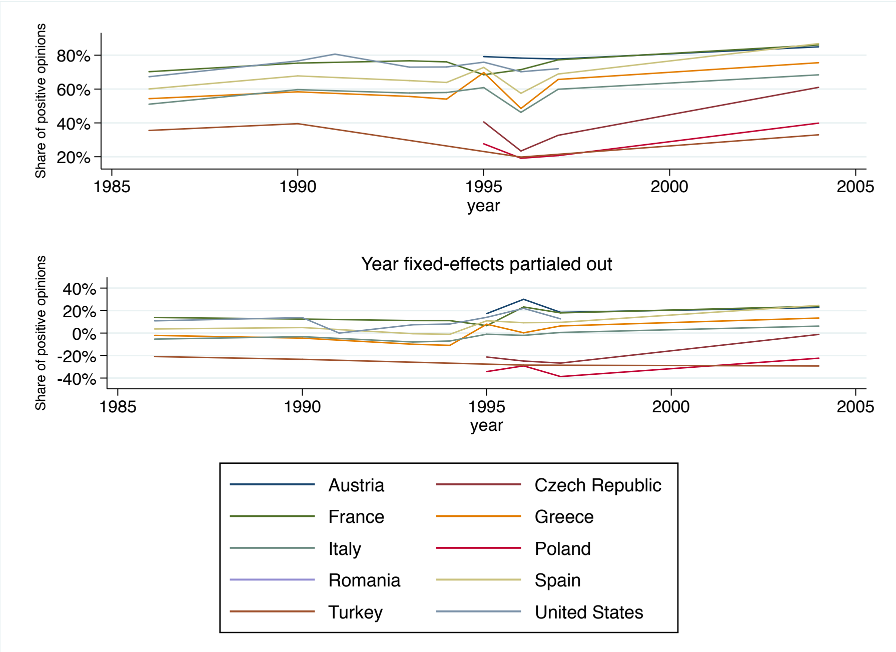
Source: Desktop Companion to the German Socio-Economic Panel (GSOEP).

Table B9: Definition and source of main variables

Variable	Definition	Source
Trust	Mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin.	Eurobarometer and EES
Trust(region)	Mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin and by German region.	Eurobarometer and EES
Discrimination	Categorical variable with three groups: it is equal to zero if the respondent never felt disadvantaged due to his origin in the year of the survey, equal to one if the respondent rarely felt disadvantaged due to his origin, and equal to two if the respondent often felt disadvantaged due to his origin.	German Socio Economic Panel
German identity	Categorical variable with five groups: feel completely German, mostly German, in some respects, barely, not at all.	German Socio Economic Panel
Age	Categorical variable with four groups: younger than 25, between 25 and 34, between 35 and 44 and above 44 years old	German Socio Economic Panel
Female	Binary variable equal to one if respondents is a woman and 0 otherwise.	German Socio Economic Panel
Married	Binary variable equal to zero if respondents is single (this category includes divorced or widowed individuals) and one if the respondent is married.	German Socio Economic Panel
Number of children	Categorical variable with five groups: no child present in the household, One child present, two children present, three children present, four children or more present in the household.	German Socio Economic Panel
Education	Categorical variable with three groups: low ISCED, middle ISCED, and high ISCED.	German Socio Economic Panel
Years since immigration	Categorical variable with seven groups: less than 5 years, between 5 and 9 years, between 10 and 14 years,	German Socio Economic Panel
	between 15 and 19 years, between 20 and 24 years, between 25 and 29 years, and 30 years and more.	
Assistance	Binary variable equal to one if respondents received (public) social assistance in the survey year and zero otherwise.	German Socio Economic Panel
Immigration Rate	Annual stock of regional population by nationality over all the residents of the region in the year. Available for all the Lander from 1991 to 2012.	German Federal Statistical Office (2016)
Stock immigrants	Annual stock of regional population by nationality. Available for all the Lander from 1984 to 2012.	German Federal Statistical Office (2016)
Unemployment rates	Monthly stock of regional unemployed population over all the active population of the region in the year.	German Federal Employment Agency (2016)

Source: Author's elaboration.

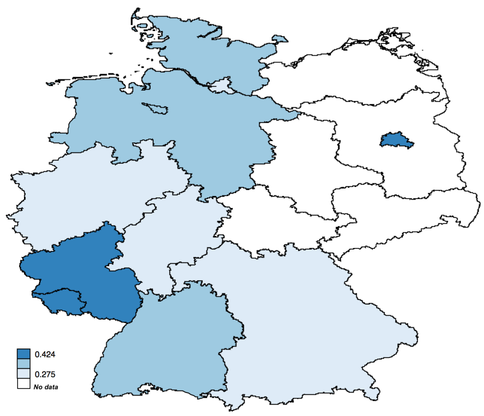
Figure B5: Share of Germans trusting citizens from other countries



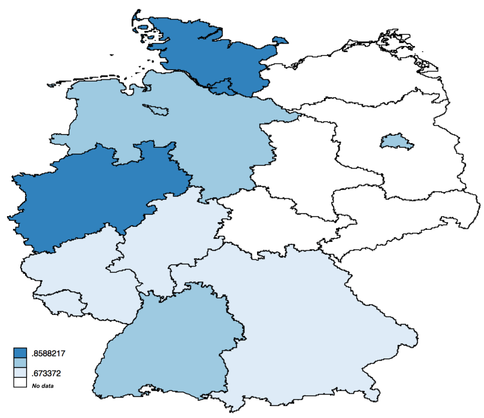
Source: Author's elaboration on Eurobarometer and European Election Survey data. The sample includes the top ten countries of origin over the period of analysis. Share of Germans trusting citizens from other countries is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin.

Figure B6: Observable characteristics, by region and origin-specific level of trust

Turkish



Austrian



Source: Author's elaboration on Eurobarometer and European Election Survey data. Trust(region) is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin and by German regions.

Table B10: Descriptives statistics by origin country.

Variable	Turkey		Greece		Italy		Spain		Austria	
	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.
Trust	0.320	0.000	0.596	0.000	0.558	0.000	0.662	0.000	0.800	0.000
Trust(region)	0.311	0.039	0.569	0.023	0.556	0.034	0.651	0.035	0.758	0.019
Age	1.541	0.999	2.183	0.889	1.925	1.016	2.081	1.013	2.426	0.870
Female	0.381	0.486	0.451	0.498	0.401	0.490	0.484	0.500	0.859	0.349
Nb. child	1.582	1.218	0.990	1.077	0.906	0.979	0.656	0.967	0.351	0.705
Education	0.597	0.604	0.624	0.747	0.580	0.606	0.544	0.705	0.777	0.416
Married	0.827	0.379	0.854	0.353	0.748	0.434	0.692	0.462	0.742	0.437
Years since migration	3.172	1.692	4.420	1.497	3.983	1.711	4.230	1.413	4.836	1.363
Assistance	0.120	0.325	0.103	0.304	0.098	0.297	0.073	0.261	0.098	0.297

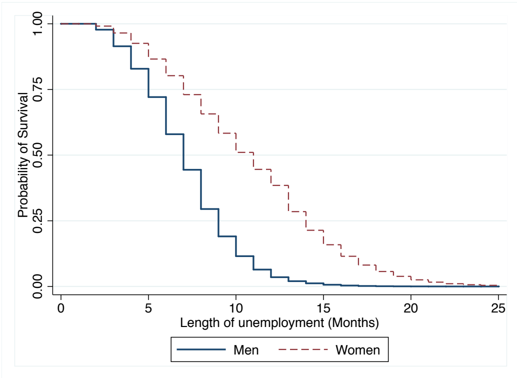
Variable	France		United Kingdom		United States		Romania		Polish	
	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.
Trust	0.689	0.000	0.565	0.000	0.685	0.000	0.244	0.000	0.268	0.000
Trust(region)	0.688	0.067	0.571	0.019	0.684	0.027	0.291	0.089	0.225	0.074
Age	1.709	0.869	2.636	0.482	1.948	0.902	2.121	0.943	2.015	0.916
Female	0.841	0.366	0.565	0.496	0.664	0.473	0.645	0.479	0.595	0.491
Nb. child	0.517	0.751	1.054	0.734	1.097	1.210	0.544	0.819	0.946	1.008
Education	1.237	0.656	1.123	0.822	1.379	0.503	1.014	0.539	1.228	0.593
Married	0.660	0.474	0.755	0.431	0.652	0.477	0.645	0.479	0.747	0.435
Years since migration	3.207	1.982	4.412	1.254	3.900	1.831	2.571	1.449	2.743	1.513
Assistance	0.121	0.326	0.082	0.275	0.099	0.299	0.157	0.363	0.150	0.357

Variable	Hungary		Portugal		Czech Republic		Russia		Netherlands	
	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.	Mean	Std.D.
Trust	0.517	0.000	0.597	0.000	0.394	0.000	0.347	0.000	0.711	0.000
Trust(region)	0.481	0.051	0.592	0.017	0.316	0.046	0.358	0.026	0.696	0.034
Age	2.284	0.734	2.466	0.905	2.273	0.894	2.157	0.939	2.140	0.866
Female	0.798	0.402	0.478	0.500	0.945	0.228	0.489	0.500	0.800	0.400
Nb. child	0.753	0.832	0.849	1.132	0.223	0.481	0.902	1.194	0.906	1.014
Education	1.541	0.612	0.231	0.422	0.945	0.375	1.301	0.651	1.492	0.640
Married	0.367	0.482	0.781	0.414	0.453	0.498	0.756	0.430	0.468	0.499
Years since migration	3.271	1.793	4.228	2.021	4.403	1.903	1.859	1.287	3.173	1.737
Assistance	0.137	0.344	0.389	0.488	0.235	0.425	0.230	0.421	0.167	0.373

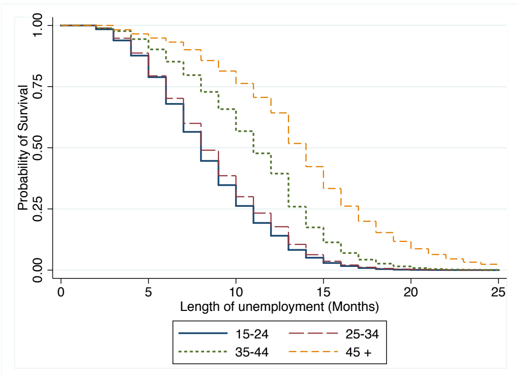
Source: Author's elaboration on GSOEP panel data. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Trust(region) is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin and by German regions. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

Figure B7: Kaplan-Meier survival estimates for covariates 1

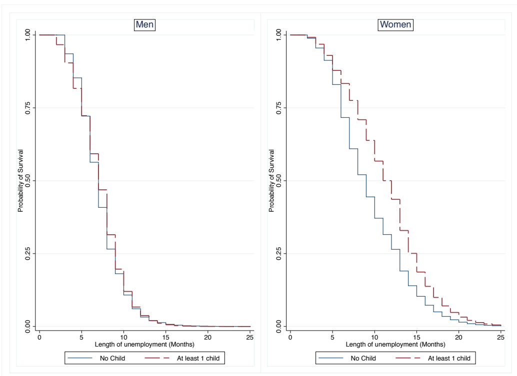
By sex



By age



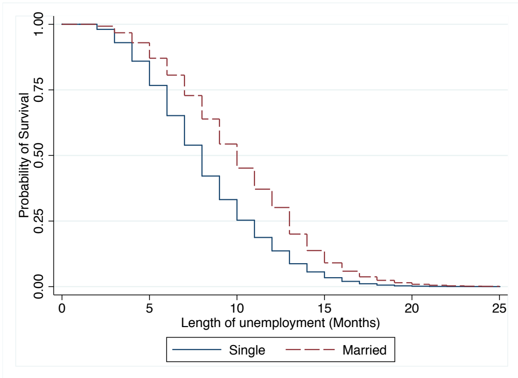
By sex and number of children



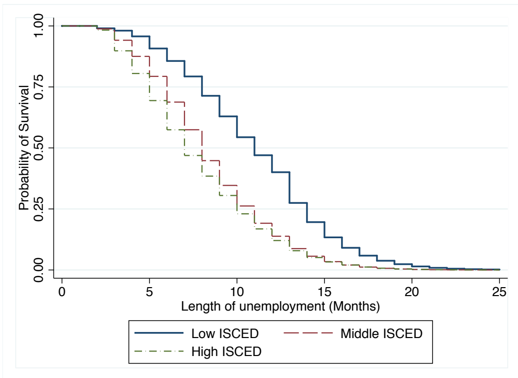
By marital status

Source: Author's elaboration on GSOEP panel data.

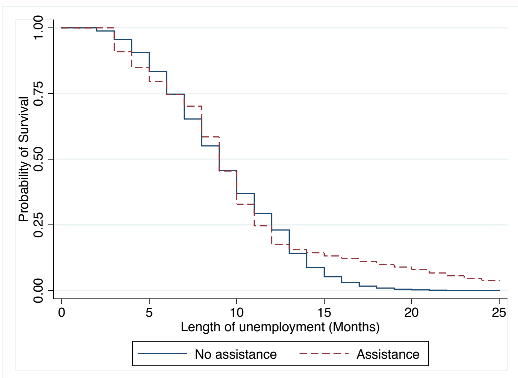
Figure B8: Kaplan-Meier survival estimates for covariates 2



By education level



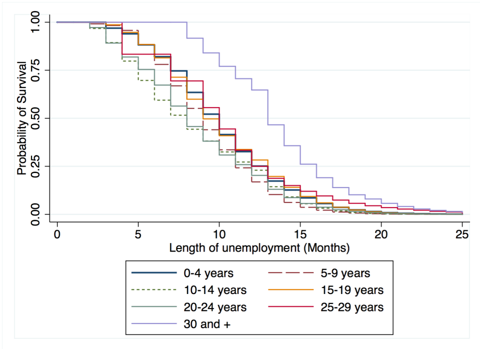
By unemployment benefit



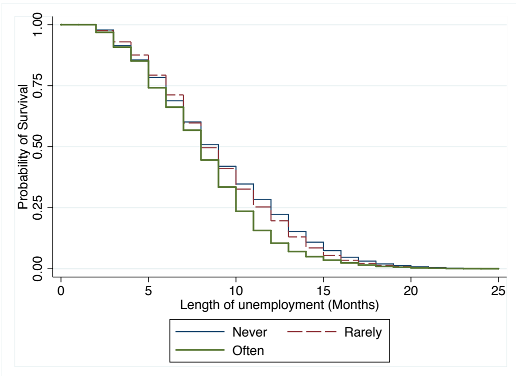
By years since migration

Source: Author's elaboration on GSOEP panel data.

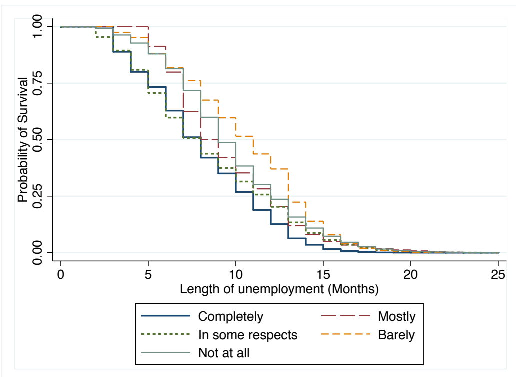
Figure B9: Kaplan-Meier survival estimates for covariates 3



By perceived discrimination



By German identity



Source: Author's elaboration on GSOEP panel data.

Table B11: Descriptives statistics (Perceived discrimination)

Variable	Mean	Std.D.	Min	Max	Log-rank Test
Discrimination	0.553	0.632	0	2	241.450***
German identity	2.096	1.300	0	4	119.990***
Age	2.016	0.923	0	3	2105.230***
Female	0.472	0.499	0	1	216.340***
Nb. Children	1.095	1.162	0	4	693.350***
Education	0.891	0.682	0	2	1145.860***
Married	0.781	0.414	0	1	62.030***
Years since migration	3.411	1.891	0	6	636.350***
Assistance	0.196	0.397	0	1	20729.750***

Source: Author's elaboration on GSOEP panel data. Discrimination refers to the question *"how often have you experienced disadvantages in the last two years because of your origins?"*. The variable takes the value of 0 if the response is *never*, 1 if the response is *seldom*, and 2 if the response is *often*. German identity refers to the question *"How much do you feel like a German?"*. German identity takes the value of 0 if the response is *completely*, 1 if the response is *mostly*, 2 if the response is *in some respects*, 3 if the response is *barely* and 4 if the response is *not at all*. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

Table B12: Origin countries of immigrants (Perceived discrimination).

Country of origin	Indiv.	Obs.	Perc.	Survival time		
				25%	50%	75%
Turkey	341	22179	29.29	8	10	13
Ex-Yugoslavia	187	8952	11.82	4	7	10
Poland	162	8538	11.26	5	8	10
Kazakhstan	139	6110	8.07	6	9	13
Russia	137	7099	9.38	9	11	14
Italy	118	7561	9.99	12	15	20
Greece	74	4514	5.96	15	15	16
Romania	45	2699	3.56	12	14	16
Ukraine	26	687	0.91	7	11	14
Spain	22	1079	1.42	13	13	14
Bosnia	13	602	0.80	4	5	7
Austria	12	588	0.78	8	8	9
Hungary	12	516	0.68	22	23	25
France	10	794	1.05	2	2	3
United States	10	731	0.97	3	3	3
Croatia	10	454	0.60	2	3	5
Kyrgyzstan	10	414	0.55	3	3	3
Netherlands	9	549	0.73	25	25	26
Portugal	9	266	0.35	13	14	16
Albania	8	418	0.55	32	33	35
Czech Republic	8	365	0.48	37	38	41
United Kingdom	7	426	0.56	13	13	14
Iran	7	192	0.25	16	17	19
Total	1376	75733	100.000	6	8	12

Source: Author's elaboration on GSOEP panel data, Eurobarometer and European Election Survey data. Survival time is the survival time to the first failure. It has to be interpreted as the number of months which have been necessary to reach that x% of the unemployed population has found a job.

Table B13: Correlation between *Trust* and individual labor earnings

	(1) Pearson correlation Earnings (log)	(2) Pearson correlation Earnings (log)	(3) Pooled-OLS Earnings (log)	(4) Pooled OLS Earnings (log)
Trust	0.074***		0.309***	
	(0.000)		(0.001)	
Trust(region)		0.067***		0.267***
		(0.000)		(0.002)
Individuals			3188	3188
Observations			23250	23250
R-squared			0.300	0.300

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer and European Election Survey data. Earnings represents the annual labor earnings of individuals 16 years of age and older. Labor earnings include wages and salary from all employment including training, primary and secondary jobs, and self-employment, plus income from bonuses, over- time, and profit-sharing. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Columns (3) and (4) include a full vector of controls with age, sex, number of children, an interaction between the number of children and sex, education, Married, years since migration, a dummy for whether the individuals receive any assistance and year and regional fixed-effects.

Table B14: Cross-correlations between *Trust*, genetic and cultural distances

	Trust	Genetic Distance	Religious Distance	Cultural distance
Trust	1.000			
Genetic Distance	-0.468***	1.000		
	(0.000)			
Religious Distance	-0.466***	0.888***	1.000	
	(0.000)	(0.000)		
Cultural distance	-0.773***	0.880***	0.866***	1.000
	(0.000)	(0.000)	(0.000)	

*** p<0.01, ** p<0.05, * p<0.1. Source: Author's elaboration on Eurobarometer data, European Election Survey, and data from Spolaore and Wacziarg (2009). Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Genetic distance accounts for the differences in the genetic composition between populations of two countries. Religious distance is the weighted religious similarity, corresponding to the religious proximity between two randomly chosen individuals between two countries. Cultural distance accounts for the differences, across pairs of countries, in average responses to 98 questions asked in the World Values Survey (WVS).

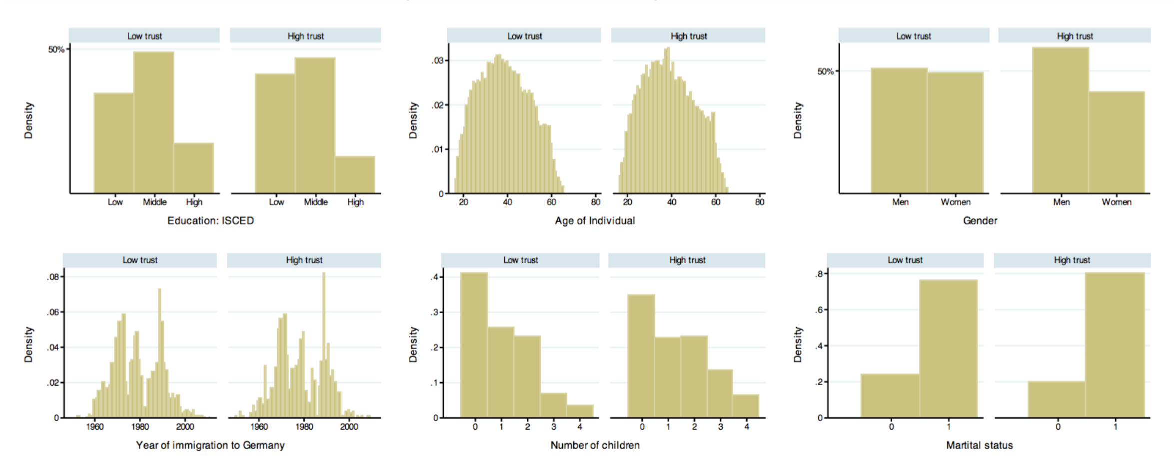
Table B15: Natives' attitudes and immigrant's unemployment duration.
Additional controls variables

	(1) Weibull Hazard		(2) Weibull Hazard		(3) Weibull Hazard		(4) Weibull Hazard		(5) Weibull Hazard		(6) Weibull Hazard	
	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.
Trust(region)	1.194***	(0.456)	1.214***	(0.462)	1.172**	(0.458)	0.552*	(0.292)	0.570*	(0.292)	0.619**	(0.298)
Speak German (ref is very good):												
Good			-0.019	(0.021)								
Fairly			-0.099***	(0.029)								
Poorly			-0.129***	(0.041)								
Not at all			-0.336**	(0.152)								
Write German (ref is very good):												
Good					-0.015	(0.023)						
Fairly					-0.054**	(0.028)						
Poorly					-0.116***	(0.034)						
Not at all					-0.213***	(0.049)						
Sectors (ref is Agriculture):									0.000		0.000	
Manufacturing									-0.172**	(0.082)	-0.130	(0.249)
Construction									-0.151*	(0.085)	-0.136	(0.220)
Services									-0.127	(0.082)	-0.207	(0.180)
Occupations (ref is Managers):												
Professionals									-0.043	(0.062)	0.094	(0.233)
Technicians									0.003	(0.054)	0.097	(0.126)
Clerical support workers									-0.021	(0.059)	0.025	(0.131)
Services and Sales workers									-0.021	(0.056)	0.077	(0.130)
Skilled Agricultural workers									0.265*	(0.152)	0.424**	(0.187)
Trade workers									-0.007	(0.056)	-0.149	(0.223)
Plant operators/Assemblers									-0.019	(0.054)	0.063	(0.129)
Elementary Occupations									-0.034	(0.055)	0.074	(0.130)
Interactions:												
Manufacturing * Professionals											-0.157	(0.278)
Manufacturing * Technicians											-0.112	(0.180)
Manufacturing * Clerical support workers											-0.116	(0.196)
Manufacturing * Services and Sales workers											-0.241	(0.238)
Manufacturing * Trade workers											0.114	(0.254)
Manufacturing * Plant operators/Assemblers											-0.123	(0.180)
Manufacturing * Elementary Occupations											-0.145	(0.182)
Construction * Professionals											-0.539	(0.458)
Construction * Technicians											-0.180	(0.152)
Construction * Clerical support workers											0.031	(0.143)
Construction * Services and Sales workers											-0.106	(0.140)
Construction * Skilled Agricultural workers											-14.743***	(1.024)
Construction * Trade workers											0.103	(0.246)
Construction * Plant operators/Assemblers											-0.012	(0.141)
Construction * Elementary Occupations											-0.171	(0.150)
Services * Professionals											-0.049	(0.200)
Services * Trade workers											0.167	(0.201)
Observations	96143		96143		96065		87585		87585		87585	
Individuals	1370		1370		1370		1154		1154		1154	
Failures	64492		64492		64492		65720		65720		65720	
Individual controls	Yes		Yes		Yes		Yes		Yes		Yes	
Seasonal fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
Origin fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	
Regional $\times$Year fixed-effects												
$ln(\rho)$	0.312***		0.306***		0.308***	(0.019)	0.161***		0.162***		0.161***	

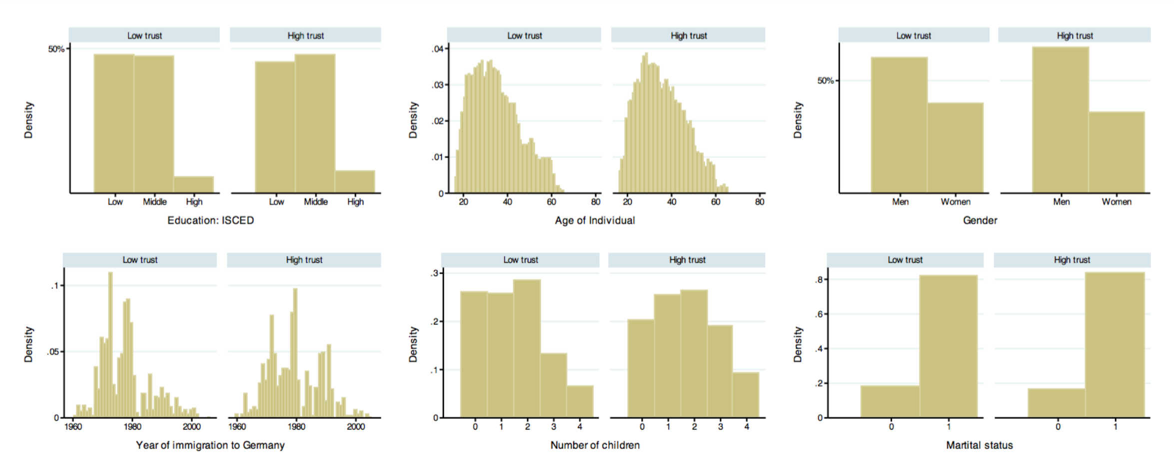
*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the individual level. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer, European Election Survey data, The German Federal Statistical Office and The German Federal Employment Agency. $ln(\rho)$ is the estimated shape parameter. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Control variables includes age category, gender, number of children, education level, marital status, years since migration, and a dummy variable equal to one if the respondent received social assistance and zero otherwise.

Figure B10: Observable characteristics, by region and origin-specific level of trust

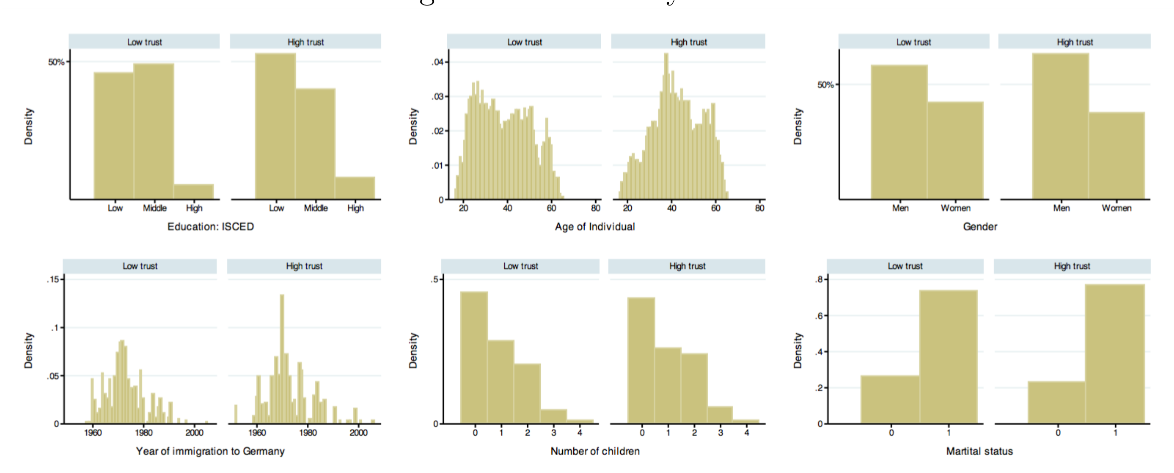
Migrants from all origins



Migrants from Italy



Migrants from Turkey



Source: Author's elaboration on GSOEP panel data. For each origin country, regions are classified into *High trust* and *low trust* if they are respectively above or below the origin-specific mean level of trust.

Table B16: Natives' attitudes and immigrant's unemployment duration. Restricted sample from 1984 to 1997.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Hazard		Hazard		Hazard		Hazard		Hazard		Hazard		Hazard	
	1984-1997		1984-1997		1984-1997		1984-1997		1984-1997		1991-1997		1991-1997	
	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.	β	Std.D.
Trust	0.428***	(0.151)												
Trust(region)			0.475***	(0.143)	1.438**	(0.699)	1.633**	(0.742)	1.721**	(0.740)	1.852**	(0.750)	2.013***	(0.759)
Stock immigrants (log)									0.134*	(0.080)				
Unemployment Rate (monthly)											-0.012	(0.020)	-0.030***	(0.010)
Age (ref<25):														
25-34 years old	-0.041	(0.042)	-0.043	(0.042)	-0.076*	(0.044)	-0.066	(0.043)	-0.067	(0.043)	-0.003	(0.069)	0.003	(0.070)
35-44 years old	-0.096*	(0.055)	-0.099*	(0.055)	-0.170***	(0.057)	-0.168***	(0.057)	-0.171***	(0.057)	-0.071	(0.081)	-0.073	(0.082)
45 and + years old	-0.425***	(0.073)	-0.429***	(0.073)	-0.520***	(0.076)	-0.508***	(0.075)	-0.511***	(0.075)	-0.461***	(0.103)	-0.451***	(0.103)
Female (ref is male)	-0.136**	(0.069)	-0.134**	(0.068)	-0.149**	(0.071)	-0.144**	(0.071)	-0.136*	(0.070)	-0.146*	(0.088)	-0.144*	(0.087)
Nb. Children (ref is no child):														
One child	0.083*	(0.050)	0.085*	(0.049)	0.100**	(0.051)	0.105**	(0.051)	0.106**	(0.051)	0.144**	(0.066)	0.139**	(0.066)
Two children	0.060	(0.055)	0.063	(0.054)	0.081	(0.057)	0.082	(0.056)	0.086	(0.057)	0.088	(0.071)	0.089	(0.072)
Three children	0.079	(0.068)	0.080	(0.067)	0.098	(0.068)	0.092	(0.068)	0.100	(0.068)	0.018	(0.087)	0.012	(0.086)
Four children and +	-0.026	(0.111)	-0.022	(0.111)	-0.018	(0.110)	-0.024	(0.106)	-0.017	(0.106)	0.071	(0.108)	0.075	(0.106)
Female * One child	-0.162*	(0.090)	-0.167*	(0.090)	-0.181**	(0.091)	-0.188**	(0.089)	-0.188**	(0.089)	-0.170	(0.112)	-0.163	(0.109)
Female * Two children	-0.068	(0.091)	-0.072	(0.090)	-0.065	(0.094)	-0.062	(0.094)	-0.068	(0.094)	-0.028	(0.113)	-0.023	(0.112)
Female * Three children	-0.264**	(0.130)	-0.265**	(0.130)	-0.238*	(0.128)	-0.256**	(0.129)	-0.263**	(0.129)	-0.164	(0.171)	-0.164	(0.172)
Female * Four children and +	-0.403	(0.275)	-0.403	(0.275)	-0.352	(0.273)	-0.342	(0.262)	-0.337	(0.264)	-0.612	(0.424)	-0.558	(0.405)
Education (ref is low ISCED):														
Middle ISCED	0.094**	(0.040)	0.096**	(0.040)	0.069*	(0.041)	0.070*	(0.041)	0.073*	(0.041)	0.048	(0.048)	0.054	(0.048)
High ISCED	0.196***	(0.058)	0.204***	(0.058)	0.144**	(0.063)	0.135**	(0.063)	0.135**	(0.062)	0.186***	(0.070)	0.186***	(0.069)
Married (ref is single)	-0.077*	(0.043)	-0.076*	(0.043)	-0.082*	(0.046)	-0.081*	(0.045)	-0.082*	(0.045)	-0.075	(0.060)	-0.070	(0.060)
Years since migration (ref<5):														
5-9 years	0.202**	(0.085)	0.206**	(0.085)	0.212**	(0.085)	0.209**	(0.083)	0.205**	(0.083)	0.138	(0.105)	0.137	(0.102)
10-24 years	0.141	(0.092)	0.145	(0.092)	0.182*	(0.095)	0.193**	(0.092)	0.183**	(0.093)	0.169	(0.125)	0.178	(0.122)
15-19 years	0.004	(0.096)	0.006	(0.096)	0.062	(0.099)	0.073	(0.096)	0.060	(0.097)	0.114	(0.128)	0.111	(0.125)
20-24 years	-0.048	(0.101)	-0.048	(0.101)	0.020	(0.104)	0.033	(0.101)	0.021	(0.102)	0.002	(0.131)	0.008	(0.128)
25-29 years	-0.118	(0.113)	-0.120	(0.112)	-0.026	(0.116)	-0.020	(0.113)	-0.029	(0.113)	-0.074	(0.137)	-0.068	(0.134)
30 and +	-0.224	(0.137)	-0.224	(0.137)	-0.130	(0.143)	-0.132	(0.139)	-0.139	(0.137)	-0.128	(0.162)	-0.121	(0.159)
Assistance	-1.478***	(0.135)	-1.478***	(0.135)	-1.481***	(0.134)	-1.475***	(0.133)	-1.472***	(0.133)	-1.448***	(0.132)	-1.451***	(0.132)
Constant	-2.690***	(0.192)	-2.718***	(0.191)	-3.317***	(0.319)	-3.126***	(0.747)	-4.521***	(1.133)	-2.719***	(0.379)	-2.466***	(0.355)
Observations	50486		50486		50486		50486		50486		28091		28091	
Individuals	1015		1015		1015		1015		1015		731		731	
Failures	30494		30494		30494		30494		30494		17764		17764	
Seasonal fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Regional fixed-effects	Yes		Yes		Yes		No		No		Yes		No	
Year fixed-effects	Yes		Yes		Yes		No		No		No		No	
Origin fixed-effects	No		No		No		Yes		Yes		No		Yes	
Origin×Year fixed-effects	No		No		Yes		No		No		Yes		No	
Regional × Year fixed-effects	No		No		No		Yes		Yes		No		Yes	
$ln(\rho)$	0.349***	(0.024)	0.349***	(0.023)	0.395***	(0.025)	0.378***	(0.024)	0.382***	(0.024)	0.320***	(0.033)	0.312***	(0.032)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the individual level. Source: Author's elaboration on SOEP panel data over 1984-2012, Eurobarometer, European Election Survey data, The German Federal Statistical Office and The German Federal Employment Agency. $ln(\rho)$ is the estimated shape parameter. Trust is the mean of the share of Germans who expressed that they trust citizens from a given country, calculated by country of origin. Age is a categorical variable with four groups: younger than 25 (0), between 25 and 34 (1), between 35 and 44 (2) and above 44 years old (3). Female is a dummy variable equal to one if the respondent is a woman and zero otherwise. Nb. children is a categorical variables with five groups: no child (0), One child (1), two children (2), three children (3) and four children (4) present in the household. Education is a categorical variable with three groups: low ISCED (0), middle ISCED (1) and high ISCED (2). Married is a dummy variable equal to one if the respondent is married and zero otherwise. Years since migration to Germany is a categorical variable with seven groups: less than 5 years (0), between 5 and 9 years (1), between 10 and 14 years (2), between 15 and 19 years (3), between 20 and 24 years (4), between 25 and 29 years (5) and 30 years or more (6). Assistance is a dummy variable equal to one if the respondent received social assistance and zero otherwise.

Chapter 4

Do Migrants Transfer Productive Knowldege Back to Their Origin Countries?

A slightly different version of this chapter is forthcoming in the *Journal of Development Studies* at http://dx.doi.org/10.1080/00220388.2017.1333109[1]

> "Accumulating productive knowledge is difficult. For the most part, it is not available in books or on the internet. It is embedded in brains and human networks." (Hausmann *et al.*, 2011, p. 7)

4.1 Introduction

The Venetian travel merchant Marco Polo spent 24 years in Asia at the end of the thirteenth century, describing his travels in *The Book of the Marvels of the World*, which allowed Europeans to discover a number of Chinese innovations–such as paper money, the use of the coal and eyeglasses–that were subsequently adopted in the West. Although information about the existence of new technologies certainly spreads across the globe more easily today than in Marco Polo's time, this needs not to translate immediately into its local adoption, which represents the main form of innovation in developing countries (World Bank, 2008). The financial returns from adopting foreign technologies are

[1] I would like to thank Simone Bertoli, Jean-Louis Combes, Frédéric Docquier, Anne Viallefont, Yves Zenou, the participants at the first Ph.D. Workshop on the Economics of Migration in Southampton, UK, and the seminar participants at CERDI for very helpful comments and fruitful discussions. I would also like to thank two anonymous reviewers for their comments and suggestions, which substantially improved the paper.

uncertain, and no legal protection is granted to an entrepreneur who succeeds in adopting a foreign technology, so that rival domestic producers can rapidly erode ensuing profits. This risk can result in the under-provision of entrepreneurial efforts required for the adoption of foreign technologies (Hausmann and Rodrik, 2003).

This paper analyses whether international migrants can contribute to fostering innovation in developing countries thus reducing the uncertainty surrounding the profitability of a local adoption of foreign technologies. Indeed, international migration can facilitate the transfer of technologies from the North to the South, by connecting high technology countries with low ones. Diasporas and emigrants are directly in touch with what is produced in developed countries and can act as scouts, exploring all the production possibilities for their origin country. They can more easily understand which technologies are suitable for local adoption and which are not, lifting the veil on the cost structure of their origin country.[2] To the best of our knowledge, the first empirical evidence of international knowledge diffusion at the country level was raised by Coe and Helpman (1995), who underlined the positive correlation between foreign R&D capital and total factor productivity. More recently, Bahar *et al.* (2014) have shown that a country is 65 percent more likely to add a new product to its export basket if a neighbouring country already exports this product. While not testing for the channels through which knowledge is spread across nations, Bahar *et al.* (2014) mention that trade, Foreign Direct Investments (FDI, thereafter) and migration certainly play a role in that process. In this paper we empirically investigate this major issue, focusing on the role of international migration. We use the Economic Complexity Index as a proxy for innovation and productive knowledge embedded in each economy. Productive knowledge encompasses the largest definition of knowledge, taking into account not only explicit but also tacit knowledge, which is both harder to transfer and the lack of which is more likely to hamper the growth of countries (Hidalgo

[2]Our paper only focuses on the effect of international emigration on migrants' origin countries. We do not investigate the effect of immigration since immigration from the North to the South only represents 6% of the total international migration (Özden *et al.*, 2011). Also, immigration stocks in developing countries requires many imputations that lower the quality of the estimates.

et al., 2007). Measuring productive knowledge allows us to capture possible knowledge spillovers between products since proximity between goods matters, and capabilities required for one product are useful in many other different productions. To address endogeneity issues, we rely on the System GMM estimator (Blundell and Bond, 1998) which allows us to deal with identification issues of our variable of interest as well as other covariates. We alternatively use internal and external instruments and the predictions of a pseudo-gravity regression that includes interactions between year dummies and the geographic distances between each destination-origin pair (Feyrer, 2009). Our results demonstrate that international migrants foster the local adoption of foreign technology in their origin countries. We also provide evidence that our main results are not driven by trade, FDI or geographical or genetic distance and that they are robust to different technological indicators.

This present paper contributes to a recent and growing body of literature that studies the diffusion of technologies across borders as a result of international human mobility. The seminal paper by Kerr (2008) shows that diasporas strongly influence the international technology diffusion. Indeed, his results point out that a ten percent increase of a given origin country's researchers residing in the United States is associated with a one percent increase in foreign output of the given origin country. His model supports the idea that scientists abroad ease the diffusion of knowledge to "technology follower's economies" and then, spur the process of imitation. Similar results are presented by Mayr and Peri (2009) with a special look at return migration and by Andersen and Dalgaard (2011) who show that the intensity of temporary movements of workers is a very good predictor of global productivity levels. Lodigiani (2008) also shows that high-skill migrants positively influence productivity levels in their origin countries as an increase in emigration rates is associated with an increase in productivity back home. In the same way, Naghavi and Strozzi (2015) concentrate their study on 34 developing countries and study the interaction between emigration and innovation performances, measured by the number of granted patents. They show that diasporas create a new source of knowledge for domestic innovators under sound intellectual property rights in the origin country. Moreover, they find that this positive inflow of knowledge overcomes the country's direct loss from emigration. At the micro level, Agrawal *et al.* (2011) use patent citation data from Indian inventions to show that diasporas abroad help inventors back home maintain access to a foreign source of knowledge and therefore spur innovation in the origin country, but only when it comes to high-quality patents. These mixed results are confirmed by Breschi *et al.* (2017) who, while demonstrating evidence of a positive diaspora effect for Asian countries, do not find significant technology transfers from the Indian diaspora to its origin country. Finally, as far as the emergence of global inventor teams is concerned, Miguélez (2016) and Kerr and Kerr (2015) study the rise in international migration which increases co-patenting and fosters the transmission of technology from developed to developing countries.

While most of the previously cited papers study technology transfers through inventor diasporas and patents, our paper uses trade and international migration data that allow us to take into account technology transfers from the least to the most highly technologically advanced exported products. Bahar and Rapoport (2016) is therefore the closest paper to our analysis since they test the hypothesis of knowledge diffusion through international migration with an analysis at the product level (1984 to 2010 UNComtrade data at the 4-digit Standard International Trade Classification). With this high level of disaggregation, they find that migration, ahead of trade and FDI, is a strong driver of the evolution of comparative advantage. An increase of ten percent in the migrant stock at destination is associated with a three percent increase in the probability to export one product for which the destination country already has a comparative advantage. Our paper enables us, using a more aggregate analysis, to capture additional indirect effects of migration on the development of new comparative advantages. Indeed, we allow migrants' origin countries to develop new comparative advantages, not only in products for which migrants' host countries have a comparative advantage, but also in products that require close productive knowledge. For instance, migrants residing in a country that exports cars (SITC 781.2) could help to promote the development of the exports by the origin country of closely related products, such as motor cycles (SITC 785.1) or motor vehicle accessories (SITC 784.3), an effect that goes beyond that identified by Bahar and Rapoport (2016).

The remainder of the paper is organized as follows. Section 4.2 discusses the data that we use, first looking at international migration data, and second focusing on the Economic Complexity Index. Next, we bring to light some stylized facts in international migration flows and technology levels. We show that recent convergence in technology is possibly associated with changes in international migration patterns. Section 4.3 presents our empirical specification and all the challenges associated with it. Section 4 outlines the baseline results and Section 5 provides some robustness checks. Finally, Section 6 concludes.

4.2 Data

This section describes the data we use in our empirical analysis. We start with international migration data and then move to the Economic Complexity Index, our index of technology.[3]

4.2.1 Emigration data

A growing body of empirical literature has recently emerged as a result of the availability of new migration data. We take advantage of these new databases and use the IAB brain-drain database developed by Brücker *et al.* (2013), which breaks down by country of origin the stocks of migrants (defined as foreign-born individuals) of 20 OECD destination countries. These destination countries are Australia, Austria, Canada, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. We rely on this dataset since migration stocks are computed on the basis of national census and population register statistics in 20 OECD countries that are among the richest migrant destinations in the world.[4] Using only 20 destination countries may be seen as problematic since we ignore half of total world

[3]Definitions, sources and statistics of all other variables can be found in the Appendix Table C5.

[4]It is worth noting that other databases provide more destination countries. However, these databases use many imputations, particularly for migration stocks in developing countries, which strongly reduces the quality of the estimates.

migration. However these 20 OECD countries receive 85 percent of high-skilled migrants worldwide. In our paper, emigration rates are defined as the share of migrants (25 years and above) in the pre-emigration population. Unfortunately, several limits emerge from migration data. It is worth noting that, although we use migration stocks, ideally researchers would be interested in flows that reflect the dynamics of the migration process. Indeed, some people that belong to a diaspora, such as retired workers for instance, do not play a huge role in technology transfers that occur to their origin country. Moreover, we know that the evolution over time of these migrant stocks reflects some demographic events that are unrelated to migration such as death or return migration, for example. Another limit is the omission of illegal migrants who are not present in census data aside from high-skilled migrants who mainly use legal channels in order to change their country of residence. We detail further some stylized facts related to recent international migration changes.

4.2.2 The Economic Complexity Index

In the long tradition of growth theories, Hausmann *et al.* (2007) have tried to understand whether countries specialising in higher productivity goods grow faster. Particularly, they have focused their attention on the sophistication of export baskets with a special look on the evolution of revealed comparative advantage (RCA thereafter). This concept, developed by Balassa (1965) , defines that a country c has a RCA in a product p only if this country exports p in larger proportions than the share of p in the world trade and is computed as:

$$RCA_{cp} \equiv \frac{x_{cp}}{\sum_c x_{cp}} / \frac{\sum_p x_{cp}}{\sum_{cp} x_{cp}} \tag{4.1}$$

where x_{cp} is the monetary value of p exported by country c. Using the notion of revealed comparative advantage, Ricardo Hausmann and his co-authors have developed two main measures of export complexity for countries. First, Hausmann *et al.* (2007) construct a proxy for what they call the productivity level associated with a country's exports (EXPY).[5] This index is based on the

[5]Hausmann *et al.* (2007) compute first PRODY an index which represents, for each product, the weighted income level associated with exporters of this product. Then, they

underlying hypothesis that higher income countries export goods with higher productivity levels. The first drawback of this indicator is that it is computed using GDP per capita and thus, presents a strong correlation with this last. The second drawback is that it does not take into account proximity between goods. Nevertheless, we know that productive knowledge required for a given product helps to the production of a similar product. These two limitations explain why Hidalgo and Hausmann (2009) recently shifted the discussion from productivity level to underlying capabilities that are required to produce a particular good. Indeed, in order to explain growth differences, the literature about export sophistication now makes reference to a new indicator: the Economic Complexity Index (ECI thereafter). In this way, Ricardo Hausmann and his co-authors write that "countries do not simply make the products and services they need. They make the ones they can" (Hausmann *et al.*, 2011, p.18). Using the ECI, the level of sophistication of a product is no longer based on the income of its exporters but on the different capabilities that it requires. Each economy can therefore be summarized through its number of available capabilities. It is worth observing that in each framework, EXPY or ECI, the authors keep the assumption that the amount of productive knowledge embedded in each country can be approximated by simply looking at national production and exports.

The ECI is built starting with a network representation of the international trade where countries are connected to tradable products, using the concept of revealed comparative advantage.[6] Countries and products (vertices) form the two independent vertex C and P of a bipartite graph $G = (C, P)$. This network can be represented through the M_{cp} matrix where each element is equal to one if the country c as a revealed comparative advantage in the product p, and zero otherwise. This matrix writes as:

$$M_{cp} \equiv \begin{cases} 1 \text{ if } RCA_{cp} \geq 1 \\ 0 \text{ otherwise} \end{cases} \tag{4.2}$$

compute EXPY as the export-weighted average of PRODY values for products exported by each country.

[6]Trade data are based on the 4-digit Standard International Trade Classifications comprising over 700 commodities.

The ECI is then computed on the basis of two different dimensions. First, Diversity wich is the number of distinctive products exported by a country. The more a country has a large panel of capabilities, the more it is able to export a diversified set of goods. Diversity is defined as:

$$Diversity_c = \boldsymbol{k}_0^c \equiv M_{cp}\mathbf{1} \tag{4.3}$$

where $\mathbf{1}$ is a column vector, whose elements are all equal to one. The second dimension in the ECI is the ubiquity. Ubiquity is a measure of sophistication that gives the number of countries c which export a given product p. A complex economy is then an economy which exports products that are exported only by itself or by a small number of countries. Ubiquity is defined as:

$$Ubiquity_p = \boldsymbol{k}_0^p \equiv \mathbf{1'}M_{cp} \tag{4.4}$$

where $\mathbf{1'}$ is the transpose of the vector $\mathbf{1}$. If we refer to the graph theory we can say that $\boldsymbol{k}_0^c$ and $\boldsymbol{k}_0^p$ are the degree of each vertex in the bipartite graph $G = (C, P)$, i.e., the number of incident paths coming from the other vertices. It is worth noticing that each dimension, namely Ubiquity and Diversity, is affected by the existence of rare capabilities. Determining if ubiquity is the result of complexity or scarcity, for instance, is then problematic. As a matter of fact, precious stones have a particularly low ubiquity since they are concentrated in few countries. However, exporting precious stones does not reveal anything about the complexity of a country since their extraction does not need complex productive knowledge. Looking at the low diversity of precious stones exporters reveals that the low ubiquity of these last comes from their scarcity. Conversely, the low ubiquity of X-ray machines does not reveal scarcity but the complex productive knowledge that are embedded in. Hausmann *et al.* (2011) use therefore an iterative process with N steps, called "the method of reflections", where the information on each dimension is used in order to correct the other. Ubiquity is computed with a corrected measure of Diversity and vice versa, such as:

$$\boldsymbol{k}_N^c \equiv \left[diag(\boldsymbol{k}_0^c)\right]^{-1} M_{cp}\boldsymbol{k}_{N-1}^p \tag{4.5}$$

$$\boldsymbol{k}_N^p \equiv [diag(\boldsymbol{k}_0^p)]^{-1} M'_{cp} \boldsymbol{k}_{N-1}^c \tag{4.6}$$

with M'_{cp} the transpose of the M_{cp} matrix and $diag(\boldsymbol{z})$ a diagonal matrix containing the elements of the vector $\boldsymbol{z}$ on its main diagonal and zero elsewhere. The method of reflections can be seen as a Markov process where each N iteration is computed using the information of the previous state $N-1$, and where $\boldsymbol{k}_0^c$ and $\boldsymbol{k}_0^p$ are the initial states. A simple way to extract all the information and the iteration processes embedded in the network, in order to obtain a proxy of the productive knowledge available in each economy, is therefore to compute the $\widetilde{M_{cc'}}$ matrix which connects each country pair. Indeed, if we insert (4.6) in (4.5) we obtain:

$$\boldsymbol{k}_N^c \equiv \widetilde{M_{cc'}} \boldsymbol{k}_{N-2}^c \tag{4.7}$$

Each entry of the $\widetilde{M_{cc'}}$ matrix represents, for a given country pair, the similarity of countries' export basket with larger weights for less ubiquitous products. This matrix is calculated as:

$$\widetilde{M_{cc'}} \equiv [diag(\boldsymbol{k}_0^c)]^{-1} M_{cp} [diag(\boldsymbol{k}_0^p)]^{-1} M'_{cp} \tag{4.8}$$

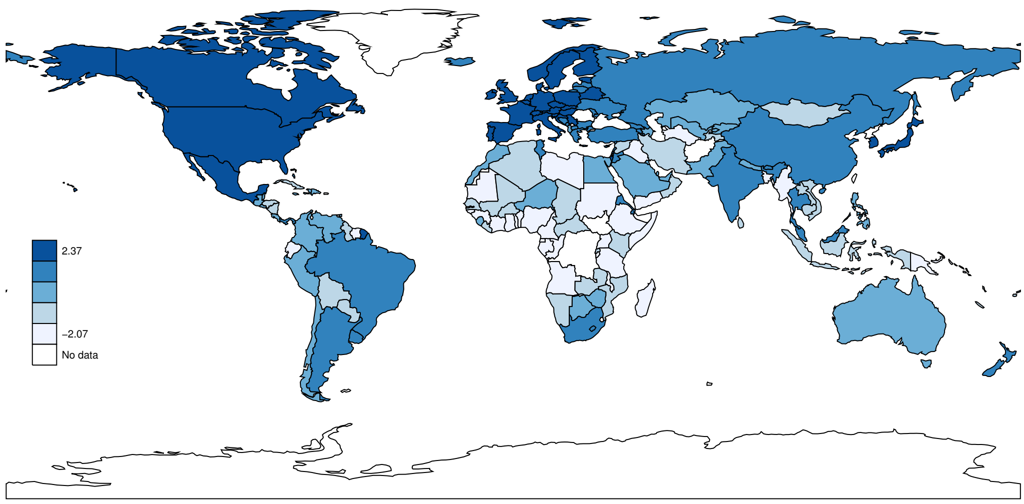
It is worth noting that, being the product of two stochastic matrices, $\widetilde{M_{cc'}}$ is also a stochastic matrix with the sum of each row equal to one. Interestingly, stochastic matrices have particular properties since their spectral radius is one and their first right eigenvector is a also a vector of ones. In order to compute the ECI, Hausmann *et al.* (2011) rely therefore on the second right eigenvector of the matrix which correspond to the second largest eigenvalue of the $\widetilde{M_{cc'}}$ matrix and which captures the largest amount of variance in the system. Finally, the ECI is the normalized eigenvector ($\vec{K}$) associated with the second largest eigenvalue of $\widetilde{M_{cc'}}$ such as:

$$ECI \equiv \frac{\vec{K} - \overline{\vec{K}}}{(\sigma_{\vec{K}})} \tag{4.9}$$

where $\overline{\vec{K}}$ representing an average, and $\sigma_{\vec{K}}$, the standard deviation of $\vec{K}$. While focusing only on the ECI in this paper, the same way we define it we can also define the Product Complexity Index. Conversely to the ECI which ranks countries, the PCI ranks products regarding their complexity, using the $\widetilde{M_{pp'}}$ matrix. This last combines information on the average diversity of countries that export a given product and on the average ubiquity of all the other products that these countries also export. Hausmann *et al.* (2011) demonstrate that the ECI captures the amount of productive knowledge that is embedded in each country. The ECI reflects both this amount and the capacity of individuals to match those different knowledge, held by distinct people, in the economy. An increase in the ECI of a given country represents therefore either new available knowledge in the economy or better abilities for people to match pre-existing knowledge (Process or organizational innovations). However, the way the ECI is defined implies that, developing a revealed comparative advantage in a product, only increases this index if the complexity of the product is larger than the complexity of the country. It means that, if Germany develops a RCA in photographic paper, it will decrease its complexity, as the PCI of photographic paper is lower than its ECI (the opposite is true for Uganda which has an ECI lower than the PCI of this product).

4.2.3 Stylized facts

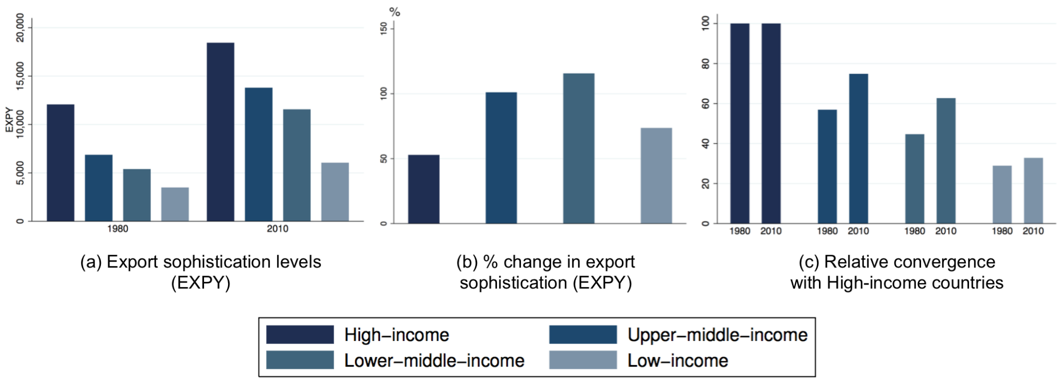
Figure 4.1: Average ECI by country from 1980 to 2010



Source: Author's elaboration on Hausmann *et al.* (2011).

Figure 4.1 depicts the average worldwide ECI between 1980 and 2010.[7] Not surprisingly, technology is unevenly distributed across the globe and sophisticated exports take place in only a few developed countries in North America and Western Europe. As a matter of fact, the correlation between the ECI and the GDP per capita is about 0.75. Figure [2(a)] confirms the gap that exists between the export sophistication of the different income groups using a simpler version of the ECI namely the EXPY. The EXPY is the export weighted average PRODY values for products exported by each country; PRODY being itself for each product, the weighted income level associated with exporters of this products (Hausmann *et al.*, 2007). However, Figures [2(b)] and [2(c)] indicate that this gap significantly narrowed over the period, leaving the scope for some convergence between the least and the most technologically advanced countries. Indeed, from 1980 to 2000, the EXPY increased twice as much in upper-middle and lower-middle income countries, than in high-income countries. Even low-income countries have started to achieve a relative convergence to the levels of the most technologically advanced countries.

Figure 4.2: Gaps in technology levels remain strong... but convergence emerges



Source: Author's elaboration on United Nations Comtrade database.

Over the same period from 1960 to 2010, international migration from the South to the North increased from 16 to 60 million people (Özden *et al.*, 2011).

[7] Post Soviet states have particularly strong average ECI here since they only present data between 1992 and 2010.

These migration flows of people have not only increased but also changed in their composition. Indeed, the number of high-skilled migrants residing in OECD countries increased by 70 percent during the 1990s, as opposed to 30 percent for low-skilled migrants (Docquier and Rapoport, 2012). Larger and more educated diasporas therefore represent greater opportunities for developing countries that would like to adopt the technologies already developed in high-income countries. These opportunities are even greater since half of all international migrants are concentrated in only 10 countries which are, for the most part, leaders in technology. This paper attempts therefore to investigate the link between the increase in migration stocks from the South to OECD countries and the convergence in technology observed over the last decades.

4.3 Empirical analysis

Our aim is to investigate whether migrants transfer productive knowledge back to their origin countries. We follow Hausmann *et al.* (2011), restricting our analysis only to countries with a population above 1.20 million between 2008 and 2010.[8] Our sample, therefore, includes 120 countries between 1980 and 2010 in five-year intervals.[9]

4.3.1 Benchmark specification

Our benchmark specification links the ECI in migrants' home countries with its values in foreign countries. Following Lodigiani and Salomone (2015) who study the transfer of gender norms related to political participation at the international level, our estimated equation is written as follows:[10]

[8]As with Hausmann *et al.* (2011), we assume that it is impossible to infer on the export structure of countries that are too small.

[9]Due to the lack of observations, the panel is unbalanced. Table C7 in the Appendix reports the number of observations for each country in the sample.

[10]Our paper borrows its main specifications from the "transfer of norms" literature. Since the seminal paper by Spilimbergo (2009) who studied the diffusion of democracy through international students, norms transferred by international migrants refer to politics (Docquier *et al.*, 2016; Barsbai *et al.*, 2017), fertily rates (Beine *et al.*, 2013; Bertoli and Marchetta, 2015) or gender norms for instance (Tuccio and Wahba, 2016; Lodigiani and Salomone, 2015).

$$ECI_{i,t} = \beta_1 \; ECI_{i,t-5} + \beta_2 \; \overline{ECI}_{i,t-5} + \boldsymbol{\lambda}'\boldsymbol{X}_{i,t} + \mu_i + \eta_t + \varepsilon_{i,t} \qquad (4.10)$$

where the dependent variable $ECI_{i,t}$ is the Economic Complexity Index in origin country i at time t and $ECI_{i,t-5}$ is its lagged value of five years. The ECI in our sample ranges from -2.78 to 2.58. $\overline{ECI}_{i,t-5}$ is our variable of interest. It is the weighted average value of the ECI in migrants' destination countries and thus captures the amount of technology to which a each given country is exposed to through its diasporas. It is computed as:

$$\overline{ECI}_{i,t-5} = \boldsymbol{w}'_{t-5}\boldsymbol{ECI}_{j,t-5} \qquad (4.11)$$

The weight vector $\boldsymbol{w}'_{t-5}$ in our benchmark specification is the emigration rates from i to j.[11] Emigration rates are the stock of migrants from origin country i residing in the destination country j, overall the total pre-emigration population of i. It is worth noting that the weight for the origin country itself is zero. $\boldsymbol{ECI}_{j,t-5}$ is the vector of complexity in the 20 OECD destination countries. By construction, our variable of interest increases, either if the ECI in any destination country increases or if the emigration rate increases to destinations for which the ECI is higher than the average ECI of the 20 OECD destination countries, or both. The average $\overline{ECI}_{i,t-5}$ in our sample equals 0.06 with a standard deviation of 0.09. Our specification allows for technology transfers even if the amount of productive knowledge at destination is lower than the level of productive knowledge at home. This relies on the nature of the ECI variable which not only summarizes the amount of productive knowledge embedded in each economy but also its composition. As a result, migrants can move to a foreign country with a lower level of technology but with a completely different export basket and still transfer new productive knowledge to their origin county. $\boldsymbol{X}_{i,t}$ is a vector of controls described in Subsection 4.3.2 and μ_i and η_t are country and time fixed effects, respectively, whereas $\varepsilon_{i,t}$ is the remaining error term. Standard errors are clustered at the country level in

[11] In Section 4.5 we test for the robustness of our results using different weight vectors accounting for bilateral trade and FDI or genetic and geographic distances.

order to correct for heteroskedasticity and serial correlation.

Our choice to rely on a dynamic panel specification borrows both from theoretical and empirical requirements. First, it takes into account the persistence in technology levels evidenced in Section 4.2. Second, it allows us to distinguish between the short-run and the long-run technological transfers that occur through emigration. Indeed, it is very likely that a given diaspora that leaves its origin country at year t will not only transfer productive knowledge to its origin country in the next five years but also in the subsequent periods. In other words, technological transfers through emigration are very unlikely to fully materialise within five years. However, the use of a dynamic panel specification calls for a careful interpretation of the coefficients. Specifically, β_2, our coefficient of interest, gives us the evidence of whether international migration acts as a channel of technological transfer from destination to migrants' origin country in the short-run. Our testable hypothesis in this paper is therefore a positive and significant β_2. Long-run effects can be obtained by dividing β_2 with the adjustment rate, namely, $1 - \beta_1$.[12]

It is important to note that usual approaches such as OLS or fixed effects are not appropriated with dynamic panel estimates. Indeed, OLS estimates are upward biased since the lagged dependent variable is correlated with the individual component of the error term. Fixed effects are not even more consistent (downward-biased) since the within transformation, in the case of samples with small T and large N, creates a correlation between the error term and the lagged dependent variable (Nickell, 1981). Our analysis relies, therefore, on the System GMM estimator (Blundell and Bond, 1998) which deals with problems of endogeneity of the lagged dependent variable. The System GMM estimator takes into account both the endogeneity of the variable of interest and all the regressors using their own lags as instruments. It combines into one system the regression in differences (Arellano and Bond, 1991) and the regression in levels (Arellano and Bover, 1995). Differences equations are instrumented with instruments in levels and levels equations are instrumented with instruments in differences. Operating in differences also permits us to control for unob-

[12]Our main results are all robust to the exclusion of the lagged dependent variable. Estimates with non dynamic panel specification with fixed effects and 2SLS fixed effects are available upon request.

served heterogeneity. The overidentification test proposed by Hansen (1982) and the autocorrelation test proposed by Arellano and Bond (1991) check for the validity of the instruments.[13]

A legitimate concern that arises from our specification is the endogeneity of the variable of interest. In the next subsection we discuss both the threats to identification that prevent us from a causal interpretation of the estimated effect and our identification strategy to overcome these issues.

4.3.2 Identification strategy

Endogeneity in our context may arise from either confounding factors or reverse causality. In order to mitigate the possibility of an omitted variable bias, our estimated equation includes both country and time fixed effects that account for country- and time-specific omitted variables such as institutions, for instance. Our equation also includes a vector of controls with time-varying variables that simultaneously affect the ECI in migrants' origin countries and the weighted-average ECI in migrants' destination countries. For the origin country i, $\boldsymbol{X}_{i,t}$ contains the average level of adult's education aged 25 and older, the logarithm of the GDP at purchasing power parity per capita, the logarithm of the population aged 25 and older, the logarithm of trade openness and FDI net inflows as a share of gross domestic product. Hausmann *et al.* (2007) shows that population, education and income positively influence export sophistication. There is indeed a positive self-perpetuating cycle between income and the adoption of foreign technologies. Larger populations also induce a larger knowledge diversity (Kuznets, 1960; Simon, 1977) as they increase the probability of having innovators who foster intellectual networks and increase markets' potential and the incentives for individuals to invest in new products (Grossman and Helpman, 1991). In terms of human capital, education is recognized as a strong determinant for the adoption of new technologies (Arrow, 1962; Romer, 1990; Grossman and Helpman, 1991) thus it increases the range

[13] We chose to keep the number of instruments below the number of groups in order to remove the problem of instrument proliferation (Roodman, 2009). All the variables, excluding the lagged dependant variable and time fixed effects, are treated as endogenous and are instrumented with their own second lag.

of discoverable goods in a given economy (Hausmann *et al.*, 2007). Finally, we include trade and FDI controls since we know that these international flows can be channels for technological transfers between nations and because they are strongly correlated with migration flows. Foreign-invested firms can directly increase the quality of exports by producing higher quality products but may also foster the production of higher technology goods in domestic firms (Javorcik, 2004). However, despite these hypotheses, the literature on FDI and knowledge transfers remains inconclusive (Görg and Strobl, 2001). Regarding trade, Madsen (2007) shows that trade openness positively impacts international knowledge transmission. In addition, developing countries are more and more frequently exposed to high technology goods, particularly if they import large quantities of intermediate goods in response to the fragmentation of the world production. These intermediate products automatically imply an increase in the export sophistication when they are re-exported as finished products (Xu, 2010).

Regarding reverse causality, Beine *et al.* (2013) show the potential biases that emerge from our specification and which can create a spurious correlation between the ECI at home and the ECI at destination. First, the interdependence of countries' weighted-average ECI in migrants' destination countries generates a reflection problem. Indeed, our variable of interest is constructed such that the weighted-average ECI in a given destination country includes the complexity of every partner with which it has developed an emigration relationship and *vice versa*. This means that, at the global level, each country's ECI depends on the other country's ECI. Second, migration is not an exogenous phenomenon; the decision to migrate is not random and poverty constraints influence the location choice of migrants. More precisely, the distribution of migrants from one country across various destinations is influenced by the level of income per capita at origin. However, we know that ECI levels are highly correlated with the level of GDP per capita in origin countries. There is, therefore, a reverse causality from the ECI at origin to our variable of interest. This can be easily illustrated by comparing Burkina Faso, a developing country, 80 percent of whose migrants have migrated to Ivory Coast, a neighbouring country with a low ECI, and France, a developed country where one of the first migrants' destinations is the United States, a distant technology leader.

This paper relies on two methods to identify whether migrants transfer productive knowledge back to their origin countries. First, we rely on the System GMM estimator, which uses lags for endogeneity issues of the variable of interest and other covariates. However, a legitimate concern with the System GMM estimator arises since it only uses internal instruments to establish causation. To tackle this issue, our analysis makes use of an external instrument for the weighted-average ECI in migrants' destination countries. We rely on Feyrer (2009) who, with panel data, extends the seminal contribution of Frankel and Romer (1999). Using the predicted values from a pseudo-gravity equation, we compute a predicted weighted-average ECI in migrants' destination countries used as an instrument for our variable of interest. As in Feyrer (2009), our time-varying source of exogeneity for bilateral migration stocks comes from the inclusion of interactions between year dummies and the log of distance between each OECD destination country and each migrants' origin country. Interactions between year dummies and the log of distance in our pseudo-gravity equation capture all of the time-varying effects of distance on migration. For instance, while the decrease in transportation and communication costs, captured through year dummies is shared by all countries, its interaction with distance generates differential changes for all country pairs.[14] Our pseudo-gravity equation is written as follows:

$$\log(Stock_{i,j,t}) = \beta_t \log(Dist_{i,j}) + Bord_{i,j} + Lang_{i,j} + Colony_{i,j} + \gamma_j + \gamma_i + \gamma_t + \varepsilon_{i,j,t}, \tag{4.12}$$

where $\log(Dist_{i,j})$ is the log of the geographical distance between origin country i and destination country j, $Bord_{i,j}$ is a dummy equal to 1 if i and j share a common border, $Lang_{i,j}$ is a dummy equal to 1 if at least 9 percent of the populations of the two countries speak a common language, γ_j, γ_i, and γ_t are the destination, origin and year fixed effects.

Obviously the exogeneity of our instrument is conditional on the other covariates included in the estimated equation. Another legitimate concern in

[14] Miguelez and Moreno (2015) and Naghavi and Strozzi (2015) also rely on the same method but with different time-varying sources of exogeneity for bilateral migration stocks.

our analysis would be that our instrument is acting through other proxies of integration such as bilateral trade or FDI, for example. In Table C9 in the Appendix we provide evidence that our results combining System GMM with an external instrument are robust to the inclusion of bilateral trade, FDI or geographical distance as controls. This strongly reduces concern for violation of the exclusion restriction.[15] In order to address the issue of the large number of zeros in migration stocks, we rely on the Poisson pseudo-maximum likelihood estimator (see Santos Silva and Tenreyro, 2006) for our pseudo-gravity equation.[16] Standard errors are clustered at the country pair level.

4.3.3 Alternative specification

While our specification borrows from Lodigiani and Salomone (2015), we also show that our results are robust to the use of the seminal equation estimated by Spilimbergo (2009). This equation can be written as follows:

$$\begin{aligned} ECI_{i,t} = \beta_1\ ECI_{i,t-5} + \beta_2\ M_{i,t-5} + \beta_3 \overline{ECI}^S_{i,t-5} \\ + \beta_4\ \overline{ECI}^S_{i,t-5} * M_{i,t-5} + \boldsymbol{\lambda' X}_{i,t} \\ + \mu_i + \eta_t + \varepsilon_{i,t} \end{aligned} \tag{4.13}$$

This interaction model links the ECI at home with the emigration rate $M_{i,t-5}$ and a weighted-average ECI in migrants' destination countries $\overline{ECI}^S_{i,t-5}$. Weights in the variable of interest are no longer emigration rates, as in our first specification, but rather emigration shares. This corresponds to the number of migrants from origin country i settled in destination j, overall the total number of migrants of i. The interaction variable crosses each constitutive term, namely $\overline{ECI}^S_{i,t-5}$ and $M_{i,t-5}$. The coefficient in front of the interaction, β_4, gives us the intuition whether emigration rates and ECI at destination play

[15]One may be concerned that origin fixed effects in the pseudo-gravity equation capture institutional variables that simultaneously affect migration and the level of productive knowledge. However, origin fixed effects are already included in our main equation, which prevents the exclusion restriction to be violated by any time-invariant country-specific variable.

[16]Results of the gravity model are available in the Appendix, Table C6.

simultaneously on ECI at home.[17] In other words, this equation allows us to test whether technological transfers are more likely to occur when emigration rates are particularly high and *vice versa*. Our testable assumption is a positive and significant coefficient for the interaction variable. We are not particularly interested in the coefficients of the two constitutive terms, β_2 and β_3, since they represent marginal effects for particular values of the conditional variables. By contrast, we plot in Section 4.4 the total effect of emigration rates and the total effect of the weighted-average ECI in migrants' destination countries for different deciles of the conditional variables following Brambor *et al.* (2006). Conditional variables are emigration rates when we look at the total effect of the weighted-average ECI in migrants' destination countries and *vice versa*.[18]

4.4 Results

Table 4.1 reports our benchmark results. From col. 1 to 3, we first investigate using OLS and fixed effects whether migrants transfer productive knowledge back to their origin countries. Year fixed effects in col. 2 and 3 account for all of the time-varying variables which similarly affect the productive knowledge levels of all of the countries in our sample. Country fixed effects in col. 3 prevent our estimates from being biased due to the omission of time-invariant country-specific factors that determine their own levels of complexity. The three estimates support the use of a dynamic panel specification since the lagged dependent variable is positive and highly significant. This highlights a strong persistence in the ECI of countries which has to be taken into account in our analysis. Regarding our variable of interest, $\overline{ECI}_{i,t-5}$ is always positive and significant at the 5% level. As far as the other covariates are concerned,

[17] We can demonstrate that $\overline{ECI}^{S}_{i,t-5}$ is a transformation of our previous variable of interest $\overline{ECI}_{i,t-5}$ such as: $\overline{ECI}_{i,t-5} = \overline{ECI}^{S}_{i,t-5} * M_{i,t-5}$. However, it is important to note that the specification following Spilimbergo (2009) suffers from collinearity problems inherent to interaction models.

[18] One may be concerned that our sample at destination always contains the same 20 OECD destination countries. However, the sample of destination countries is observed between 1980 and 2010 in five-year intervals, resulting in 140 different values for the ECI over the period. Figure C4 in the Appendix depicts the residual variability of the ECI in destination countries when country and year fixed effects are partialled out.

human capital, income and population are positively correlated with the levels of productive knowledge in migrants' home countries. However, while trade openness seems to act as a channel of technological transfer in col. 1 and 2, this is no longer the case when country fixed effects are included. Finally, FDI net inflows have no significant impact on the ECI. This underlines the mixed results in the literature on the effect of FDI on export sophistication.

While these initial estimates suggest that variations in the weighted-average ECI in migrants' destination countries are positively associated with variations in the ECI at home, it is worth remembering form Section 4.3 that OLS and fixed effects are biased with panel dynamic estimates. Thus, we turn to System GMM estimates from col. 4 to 7 that correct for the endogeneity of the lagged dependent variable and other covariates.[19] The coefficient of the lagged dependent variable β_1 with the System GMM estimator ranges between 0.690 and 0.756, and is always significant at the one percent level. This means that it takes between 11 and 14 years after a shock (2.236 and 2.841 periods of five years) before closing half of the gap with the long-run level of the ECI (see Vu (2013) for the interpretation of beta-convergence effects). As expected, this coefficient ranges between the lower (0.165) and the upper bounds (0.778) given by the previous OLS and fixed effects estimates.

Column 4 reports the results of our preferred specification. The short-run coefficient for the variable of interest is positive and significant. This indicates that migrants transfer technology to their origin countries in the next five years after they have left their origin country.[20] As far as the magnitude of the effect is concerned, a one unit increase of the weighted-average ECI in migrants' destination countries is associated with a 0.940 increase of the ECI at origin in the five years following emigration. Such an interpretation of the

[19] We check for the validity of the estimator in the last rows of Table 4.1. In every columns, we always reject the null hypothesis of first-order serial correlation and do not reject the null hypothesis of no second-order correlation in the residuals. Moreover, the Hansen's J test confirms the overall validity of the instruments. It is important to note that our p-values for this last test are particularly low. It is a great support for the validity of System GMM estimates since Roodman (2009, p. 129) recalls that Hansen test p-values way above 0.25 have to be seen as potential signs of trouble.

[20] In the Appendix Table C10, we provide evidence that our baseline result is robust to alternative lag structures either for the lagged dependent variables and other control variables.

Table 4.1: The effect of ECI in migrants' destination countries on ECI at home
Benchmark estimates (Dep= $ECI_{i,t}$)

	(1) Pooled OLS	(2) Fixed effects	(3) Fixed effects	(4) System GMM	(5) System GMM	(6) System GMM Ext. Instrument	(7) System GMM
$ECI_{i,t-5}$	0.778***	0.752***	0.165***	0.724***	0.716***	0.756***	0.690***
	(0.038)	(0.042)	(0.058)	(0.069)	(0.068)	(0.072)	(0.067)
$\overline{ECI}_{i,t-5}$	0.439**	0.446**	2.042**	0.940**		0.614**	
	(0.182)	(0.201)	(0.809)	(0.416)		(0.284)	
$\overline{ECI}^{INI}_{i,t-5}$					0.821*		
					(0.419)		
$\overline{ECI}^{S}_{i,t-5}$							-0.001
							(0.135)
$M_{i,t-5}$							-1.991
							(1.657)
$\overline{ECI}^{S}_{i,t-5} \times M_{i,t-5}$							2.102*
							(1.083)
$log(GDP_{i,t})$	0.133***	0.140***	-0.123	0.047	0.049	0.045	0.082*
	(0.026)	(0.027)	(0.078)	(0.052)	(0.047)	(0.053)	(0.044)
$log(Pop_{i,t})$	0.069***	0.083***	0.390**	0.108***	0.099***	0.084**	0.118***
	(0.015)	(0.017)	(0.156)	(0.035)	(0.033)	(0.036)	(0.038)
$Hum_{i,t}$	0.012*	0.021***	0.089**	0.066***	0.070***	0.062***	0.064***
	(0.007)	(0.008)	(0.036)	(0.024)	(0.023)	(0.023)	(0.021)
$log(Trade_{i,t})$	0.070*	0.104**	0.107	-0.092	-0.046	-0.100	-0.101
	(0.040)	(0.045)	(0.108)	(0.083)	(0.077)	(0.090)	(0.093)
$FDI_{i,t}$	-0.004	-0.003	-0.001	0.010	0.011	0.009	0.007
	(0.003)	(0.003)	(0.003)	(0.006)	(0.007)	(0.007)	(0.007)
Observations	600	600	600	600	600	600	600
Nb. countries	120	120	120	120	120	120	120
Nb. instruments				84	84	77	100
R-squared	0.908	0.911	0.154				
Country fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)				0.000	0.000	0.000	0.000
AR(2)				0.449	0.493	0.431	0.434
Hansen J (p-value)				0.237	0.236	0.278	0.170

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level (*** p<0.01, ** p<0.05, * p<0.1). From equations (4) to (7) the lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. In columns (4), (5) and (7) all the variables are treated as endogenous and instrumented with their own second lag. In column (6), $\overline{ECI}_{i,t-5}$ is instrumented using its predicted value obtained from the pseudo-gravity model described in the Appendix Table C6.

coefficient is not intuitive since a one-unit increase of the variable of interest corresponds to two times the highest observed value in the sample and thus is very unlikely to occur for any country. A more self-explanatory interpretation of this coefficient requires, therefore, that we compute the variation of the ECI at home for a one standard deviation of the variable of interest. Using a simple transformation we find that an increase of one standard deviation of the weighted-average ECI in migrants' destination countries, that is, an increase of 0.088 units, is associated with an increase of 0.083 units of the ECI at origin. Mexico, the majority of whose emigration is to the United States (more than 98% in 2010), represents a suitable case to illustrate the economic implications of the estimated coefficient. It allows us to easily quantify the contribution of migrants to the evolution of the ECI in their origin country. Indeed, if no emigrants had left Mexico for the US in 2005, then its weighted-average ECI at destination would have been equal to zero rather than 0.210. This implies that Mexico's ECI would have been 0.197 units lower than its observed value in 2010. This decrease corresponds to a drop of three places in the 2010 ECI overall ranking. In the long-run, using the adjustment rate in the dynamic panel specification, the ECI of Mexico would be 0.715 units lower, which corresponds to a drop of 25 places.

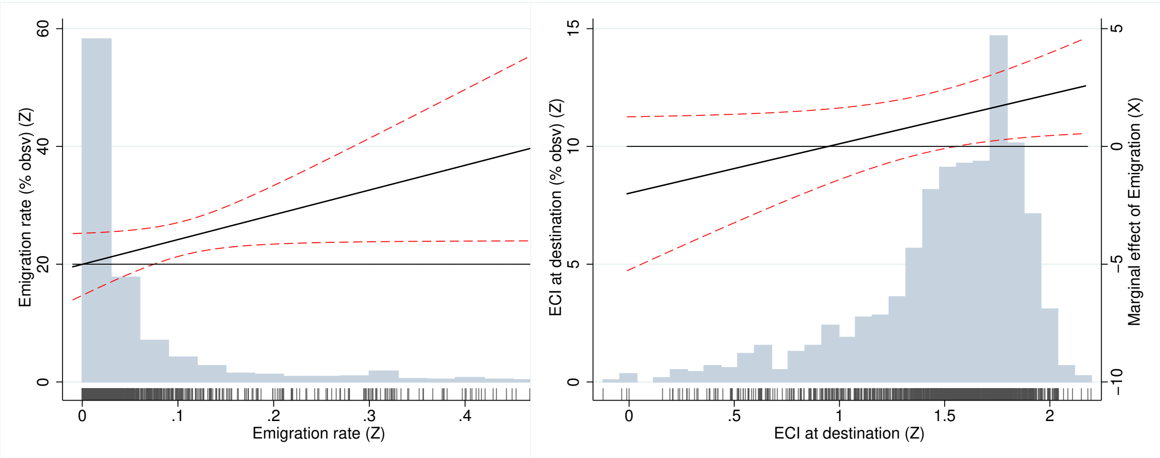
Another difficulty of interpreting ECI variations arises from the fact that the index is computed independently for each year of our period of analysis, between 1980 and 2010. This implies that changes in the world trade structure may affect the ECI trend, making the interpretation of the indicator more difficult over time. In order to control for the possible bias related to the use of the ECI in a dynamic framework, Jarreau and Poncet (2012) and Poncet and Starosta de Waldemar (2013) propose looking at the sensitivity of the results using a time-invariant measure of the ECI.[21] In the same way, we compute our variable of interest no longer using the ECI at year t but rather its initial value in 1980. This allows us to keep the amount of productive knowledge in our destination countries constant over the period. It also means that the variability over time of the latter variable now only comes from the weights vector $\boldsymbol{w}'_{t-5}$. Results are reported in col. 5. The coefficient for $\overline{ECI}^{INI}_{i,t-5}$ remains positive and significant, which underlines that removing the dynamic of productive knowledge accumulation at destination does not affect therefore our main result. The amount of technology in migrants' destination countries still positively influences the amount of productive knowledge in migrants'

[21] In these two papers the authors test the effect of export sophistication on growth, first using a time invariant measure of export sophistication and then, exploiting the variation of export sophistication over time.

origin countries.

Concerning endogeneity of the variable of interest, col. 6 reports the results of the System GMM estimator combined with an external instrument. It is worth noting that, while our variable of interest is instrumented using the predicted bilateral migration stocks obtained from our pseudo-gravity model, following Feyrer (2009), we still use internal instruments in order to correct for the endogeneity of other covariates. Despite a slight decrease in the coefficient, we still observe a positive and significant relationship between the weighted-average ECI in migrants' destination countries and the ECI at home. In fact, the coefficient is not statistically different from our benchmark coefficient reported in col. 4.

Figure 4.3: Total effect of ECI at destination and emigration rates



Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: The solid line represents the marginal effect of X conditional on all values of the modifying variable Z. The histogram indicates the percentage of observations of the modifying variable and each bar on the rug plot represents one observation for this one. The dashed lines are the upper and the lower bound of the 95 percent confidence interval, respectively.

Finally, col. 7 reports the results of the alternative specification that follows the seminal equation estimated by Spilimbergo (2009). Results for this interaction model are depicted in Figure 4.3 and shows that the ECI in migrants' destination countries has no effect on the ECI at origin when emigration rates are at their lowest levels. However, when emigration rates increase, we observe a positive and significant effect of the technology at destination on the technology in origin countries. This effect gets stronger as emigration rates increase. In the same way, emigration rates begin to have a positive and significant effect on the ECI at home only for higher levels of the technology at destination, i.e, the more sophisticated the foreign technology, the stronger the effect. The seminal specification of Spilimbergo (2009) thus supports the hypothesis of technological transfers through international emigration but highlights that this diffusion is more likely to occur with high emigration rates and high levels of technology at destination.

4.5 Robustness checks

This section first investigates whether our previous results are robust to the introduction of additional controls and sub-samples and then explores some transmission channels.

4.5.1 Sub-samples and additional control variables

Table 4.2 reports our results with sub-samples and additional control variables. Col. 1 replicates our benchmark specification for the sake of comparison. In col. 2 and 3 we split our baseline sample between developing and developed countries. In both cases the weighted-average ECI in migrants' destination countries remains significant at the five percent level. While this indicates that our results are not only driven by high-income countries, the higher coefficient for high-income countries suggests that knowledge circulation is stronger among OECD countries than among other.

As a second robustness check and to dismiss alternative explanations, we introduce a new set of weighted-average ECI variables which take into account other proxies of integration and distances between migrants' destination and home countries. More precisely, we successively replace the weight vector $\boldsymbol{w}'_{t-5}$ in our variable of interest with bilateral trade and FDI and the inverse of geographical and genetic distances.[22] It is worth noticing that this exercise is

[22]This new set of estimates suffers from data constraints which reduces the size of the

Table 4.2: Sub-samples and additional controls
System GMM (Dep= $ECI_{i,t}$)

	(1) All 1980-2010	(2) Non OECD 1980-2010	(3) OECD 1980-2010	(4) All 1980-2010	(5) All 1980-2010	(6) All 1980-2010	(7) All 1980-2010	(8) All 1985-2010
$ECI_{i,t-5}$	0.724***	0.549***	0.681***	0.647***	0.694***	0.714***	0.580***	0.717***
	(0.069)	(0.076)	(0.143)	(0.088)	(0.065)	(0.077)	(0.085)	(0.055)
$\overline{ECI}_{i,t-5}$	0.940**	0.865**	1.584**	0.844**	0.937**	0.992**	0.969***	0.586*
	(0.416)	(0.366)	(0.746)	(0.374)	(0.377)	(0.408)	(0.364)	(0.305)
$\overline{ECI}_{i,t-5}^{DIS}$				17.634***			21.451***	14.945***
				(5.995)			(6.427)	(5.571)
$\overline{ECI}_{i,t-5}^{GEN}$					0.110**		0.038	-0.002
					(0.052)		(0.053)	(0.043)
$\overline{ECI}_{i,t-5}^{IMP}$						-0.131	-0.180	0.076
						(0.101)	(0.126)	(0.127)
$\overline{ECI}_{i,t-5}^{FDI}$								-0.035
								(0.124)
$log(GDP_{i,t})$	0.047	0.098*	0.129	0.078	0.052	0.076*	0.087**	0.086**
	(0.052)	(0.057)	(0.173)	(0.050)	(0.047)	(0.045)	(0.044)	(0.038)
$log(Pop_{i,t})$	0.108***	0.053	0.196**	0.113***	0.101***	0.098***	0.114***	0.079***
	(0.035)	(0.047)	(0.095)	(0.036)	(0.035)	(0.033)	(0.035)	(0.027)
$Hum_{i,t}$	0.066***	0.073***	0.015	0.050**	0.070***	0.056**	0.059***	0.026*
	(0.024)	(0.028)	(0.042)	(0.022)	(0.024)	(0.025)	(0.022)	(0.015)
$log(Trade_{i,t})$	-0.092	-0.087	0.248	-0.130	-0.113	-0.093	-0.121	-0.028
	(0.083)	(0.120)	(0.272)	(0.084)	(0.079)	(0.086)	(0.085)	(0.075)
$FDI_{i,t}$	0.010	-0.014	-0.004	0.005	-0.002	0.005	-0.009	-0.009*
	(0.006)	(0.012)	(0.011)	(0.007)	(0.006)	(0.005)	(0.007)	(0.005)
Observations	600	411	189	600	598	600	598	404
Nb. countries	120	83	37	120	119	120	119	117
Nb. instruments	84	84	35	92	92	92	108	101
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.449	0.527	0.413	0.470	0.382	0.443	0.476	0.253
Hansen J (p-value)	0.237	0.405	0.100	0.210	0.261	0.193	0.198	0.580

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level (*** p<0.01, ** p<0.05, * p<0.1). The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. All the other variables are treated as endogenous and instrumented with their own second lag. Genetic distance data are not available for Yemen. FDI data had no records for Tajikistan or Turkmenistan or the year 1980.

particularity challenging given the strong correlations that exist between the different weighted-average ECI computed with different weights. As a matter of fact, Table C8 in the Appendix reports the Pearson correlations between the

sample. For instance, genetic distance data are not available for Yemen while FDI data are only available from 1985 and not available for Tajikistan and Turkmenistan.

different weighted average ECI. As expected, these correlations are particularly high even if using immigration rates (for $\overline{ECI}_{i,t-5}$) rather than immigration shares (as in $\overline{ECI}^{S}_{i,t-5}$), which mitigates the problem.

In col. 4 of Table 4.2 we first weight the ECI of foreign countries using the inverse of the bilateral great-circle distance between each destination-origin pair. Indeed, Bahar *et al.* (2014) show that knowledge diffusion decreases with geographical distance and that closer countries are more likely to share common technologies. The coefficient of the geographical distance weighted-average ECI ($\overline{ECI}^{DIS}_{i,t-5}$) is positive and highly significant which confirms that distance is crucial in technological transfers. The greater the distance between two countries, the lower the technological transfers. Interestingly, adding this new control does not significantly affect the coefficient of our variable of interest, which is still positive and significant at the five percent level despite a small decrease. We repeat the same exercise in col. 5 using the inverse of the genetic distance as a weight for the ECI in foreign countries. The rationale is given by by Spolaore and Wacziarg (2009), who demonstrate that genetic distance acts as a barrier to the diffusion of development from the world technological frontier. Genetically closer societies are more likely to exchange and to learn from each other. As the authors argue, similarities in terms of genetics facilitate the diffusion and the adoption of "complex technological and institutional innovations" (Spolaore and Wacziarg, 2009, p. 471). As for geographical distance, the coefficient of our variable of interest remains positive and strongly significant while the coefficient of $\overline{ECI}^{GEN}_{i,t-5}$ suggests that genetic distance plays a role in technological transfers.

Another legitimate concern in our analysis would be that confounding proxies of integration such as trade and FDI could prevent us from showing evidence of an effect of emigration on technological transfers. Indeed, trade openness captures the fact that, the more a country is trading, the more it is exposed to a diversified set of productive knowledge. However, while trade theories predict that openness to trade increases a country's specialization, they underline that this specialization depends on the initial patterns of comparative advantages. A country with comparative advantages in low-technology products will specialise in these kinds of goods as it opens its economy to trade and vice versa.

Col. 6 includes $\overline{ECI}^{IMP}_{i,t-5}$, a weighted average ECI in trade import partners. $\boldsymbol{w}'_{t-5}$ is the share that each OECD destination country represents in the total imports of the origin country. The coefficient of the weighted-average ECI in trade import partners is not significant. However, its negative sign might reflect the fact that importing a large share of products from high-complex economies reflects a strong inability for low ECI countries to produce high-technology goods locally. Still, our variable of interest remains robust to this new control.

Due to data limitations in FDI, we first introduce in the same equation (in col. 7) all of the previous additional controls added separately along the previous lines. Despite the inherent problems of collinearity, our main coefficient remains positive and becomes significant at the one percent level. Next, in col. 8 we add $\overline{ECI}^{FDI}_{i,t-5}$ a weighted average ECI in foreign countries where the weights are the share that each OECD destination country represents in the total FDI that the origin country i receives from its partners. It is worth noting that the estimated sample is restricted to the 1985-2010 period.[23] Migration ahead of trade and FDI is still found to be a strong channel of technology transmission between countries. Despite a slight decrease, the coefficient of $\overline{ECI}_{i,t-5}$ remains positive and significant.

4.5.2 Alternative channels of transmission

While the focus of the previous section was to change the weight vector in our variable of interest, this section investigates some transmission channels, adjusting the second component of the weighted-average ECI in migrants' destination countries, namely, the indicator of technology at destination. We replace the ECI in foreign countries using different proxies for the level of productive knowledge in migrants' destination countries. It is worth noting that due to collinearity we cannot add all of these new variables and $\overline{ECI}_{i,t-5}$ in the same regression. While this would have been a suitable way to test for some transmission channels, collinearity forces us to add the variables separately in

[23]For comparison, in the same specification as col. 7 i.e without $\overline{ECI}^{FDI}_{i,t-5}$ and over the same time span, the coefficient for $\overline{ECI}_{i,t-5}$ equals to 0.972 and is significant at the five percent level.

Table 4.3: Channels of transmission
System GMM (Dep= $ECI_{i,t}$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$ECI_{i,t-5}$	0.724***	0.725***	0.725***	0.721***	0.730***	0.720***	0.720***
	(0.069)	(0.070)	(0.070)	(0.071)	(0.069)	(0.070)	(0.066)
$\overline{ECI}_{i,t-5}$	0.940**						
	(0.416)						
$\overline{log(GDP_{i,t-5})}$		0.137*					
		(0.074)					
$\overline{log(TFP_{i,t-5})}$			0.127*				
			(0.069)				
$\overline{log(EXPY_{i,t-5})}$				0.150*			
				(0.083)			
$\overline{log(Patent_{i,t-5})}$					0.119*		
					(0.061)		
$\overline{log(RDexp_{i,t-5})}$						0.138**	
						(0.065)	
$\overline{log(RDwork_{i,t-5})}$							1.037**
							(0.404)
$log(GDP_{i,t})$	0.047	0.045	0.045	0.042	0.046	0.046	0.054
	(0.052)	(0.049)	(0.049)	(0.048)	(0.047)	(0.049)	(0.049)
$log(Pop_{i,t})$	0.108***	0.102***	0.102***	0.102***	0.098***	0.103***	0.096***
	(0.035)	(0.037)	(0.036)	(0.037)	(0.034)	(0.036)	(0.030)
$Hum_{i,t}$	0.066***	0.068***	0.068***	0.070***	0.068***	0.069***	0.066***
	(0.024)	(0.023)	(0.023)	(0.022)	(0.023)	(0.023)	(0.023)
$log(Trade_{i,t})$	-0.092	-0.049	-0.049	-0.050	-0.079	-0.050	-0.080
	(0.083)	(0.084)	(0.084)	(0.083)	(0.078)	(0.082)	(0.082)
$FDI_{i,t}$	0.010	0.010	0.010	0.010	0.009	0.010	0.007
	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)
Observations	600	600	600	600	600	600	600
Nb. countries	120	120	120	120	120	120	120
Nb. instruments	84	84	84	84	84	84	84
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.449	0.475	0.476	0.488	0.452	0.476	0.456
Hansen J (p-value)	0.237	0.228	0.230	0.227	0.337	0.243	0.302

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors are in parentheses are clustered at the country level (*** p<0.01, ** p<0.05, * p<0.1). The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. All of the other variables are treated as endogenous and instrumented with their own second lag.

different regressions, excluding our variable of interest from the latter.

Col. 1 in Table 4.3 replicates our baseline result for comparison. In col. 2 and 3 we first use the logarithm of the GDP per capita and the logarithm of the TFP as proxies for technology at destination. Indeed, wealthier countries with higher levels of productivity are more likely to present greater stocks of productive knowledge. In both cases the coefficient of the variable of interest is positive while only significant at the ten percent level. Similar results are obtained in col. 4 using the EXPY.

Since traditional economic indicators support the idea of a technological transfer through migration but only with low levels of significance we move to innovation indicators from col. 5 to 7. We sequentially use the logarithm of the number of patent applications made by residents, the logarithm of research and development expenditures (in million US dollars) and the logarithm of the number of researcher per 1,000 employed as indicators for technology in migrants' destination countries.[24] While the coefficient of the variable of interest computed using patents data is still only significant at the ten percent level, coefficients using indicators of the research sector in destination countries are significant at the five percent level. This highlights the importance of the research sector at destination when it comes to transfer of technology to migrants' home countries, as first underlined by Kerr (2008).

4.6 Conclusions

Technology has been recognized as one of the main determinants of development, while the distribution of productive knowledge is still unevenly distributed across the world and that the majority of regions significantly lag behind in terms of export sophistication.

This paper has shown that international migration acts as a transmission channel for technology from migrants' destination to origin countries. Using economic complexity, as a proxy for export sophistication and productive knowledge embedded in foreign countries, allows us to capture knowledge spillovers in the production process. Our results are robust to different estimation methods, to the introduction of different weights in the average ECI in migrants' destination countries and to different technology indicators. En-

[24] Patent data indicators are obtained from the World Development Indicators, while other innovation variables are obtained from the OECD Science, Technology and R&D statistics.

dogeneity issues have been addressed using the System-GMM estimator with internal and external instruments. Moreover, using the seminal specification proposed by Spilimbergo (2009), we have found that technological transfers are more likely to occur when the intensity of emigration is high and when technology levels in destination countries are high. Our results, therefore, support the statement that productive knowledge is deeply embedded in brains and human networks (Hausmann *et al.*, 2011). They also link the increase in international migration stocks from Southern to Northern countries with the convergence in technology levels observed over recent decades.

While our paper underlines that migrants transfer technology to their origin countries, our study is inconclusive about the total effect on emigration at origin. It is worth noting that the departure, particularly of high-skilled inhabitants, may also deter the adoption of foreign technology by taking away those most likely to engage in entrepreneurial activities. Future research is therefore needed to investigate whether the transmission of productive knowledge from diasporas is able to exceed the negative impact of the brain drain in developing countries. Analysis at the firm level should also allow for better tracking of the channels through which migrants spur the development of new comparatives advantages in their origin countries. Nevertheless, our results encourage migrants' origin countries to foster a close relationship with their diaspora and facilitate the ease of return migrants.

References

Agrawal, A., Kapur, D. McHale, J. and 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.

Andersen, T. B. and Dalgaard, C. J. (2011). Flows of people, flows of ideas, and the inequality of nations. *Journal of Economic Growth*, 16(1):1–32.

Arrow, K. (1962). The economic implications of learning by doing. *The Review of Economics Studies*, 29(3):155–173.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51.

Bahar, D., Hausmann, R., and Hidalgo, C. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics*, 92(1):111–123.

Bahar, D. and Rapoport, H. (2016). (in press). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal.*, forthcoming.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School*, 33(2):99–123.

Barsbai T., Rapoport, H., Steinmayr, A., and Trebesch, C. (2017). The effect of labor migration on the diffusion of democracy: Evidence from a former Soviet Republic. *American Economic Journal: Applied Economics.*

Beine, M., Docquier, F., and Schiff, M. (2013). International migration, transfer of norms and home country fertility. *Canadian Journal of Economics*, 46(4):1406–1430.

Bertoli, S. and Marchetta, F. (2015). Bringing it all back home–return migration and fertility choices. *World Development*, 65:27–40.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1):63–82.

Breschi, S., Lissoni, F., and Miguelez, E. (2017). Foreign-origin inventors in the USA: Testing for diaspora and brain gain effects. *Journal of Economic Geography*. doi: https://doi.org/10.1093/jeg/lbw044

Brücker, H., Capuano, S., and Marfouk, A. (2013). Education, gender and international migration: Insights from a panel-dataset 1980-2010, Mimeo.

Coe, D. T. and Helpman, E. (1995). International R&D spillovers. *European Economic Review*, 39(5):859–887.

Docquier, F., Lodigiani, E., Rapoport, H., and Schiff, M. (2016). Emigration and democracy. *Journal of Development Economics*, 120(1):209–223.

Docquier, F. and Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of Economic Literature*, 50(3):681–730.

Feyrer, J. (2009). Trade and income: Exploiting time series in geography. NBER working paper No. 14910. National Bureau of Economic Research Cambridge, MA.

Frankel, J.A., and Romer, D.H. (1999). Does trade cause growth? *American Economic Review*, 89(3):379–399.

Görg, H., Strobl, E. (2001). Multinational companies and productivity spillovers: A meta-analysis. *Economic Journal*, 111(475):723–739.

Grossman, G. and Helpman, E. (1991). Innovation and growth in the global economy. Cambridge: MIT Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hausmann, R. and Rodrik, D. (2003). Economic development as self-discovery. *Journal of Development Economics*, 72(2):603–633.

Hausmann, R. and Rodrik, D. (2005). Self-Discovery in a development strategy for El Salvador. *Economia: Journal of the Latin American and Caribbean Economic Association*, 6(1):43–102.

Hausmann, R., Hwang, J., and Rodrik, D. (2007). What you export matters. *Journal of Economic Growth*, 12(1):1–25.

Haussman, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., and Yildirim, M. A. (2011). The Atlas of Economic Complexity. Puritan Press. Cambridge MA.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after the independence. *Journal of International Economics*, 81(1):1–14.

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.

Hidalgo, C. A., Klinger, B., Barabási, A. L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482–487.

Jarreau, J. and Poncet, S. (2012). Export sophistication and economic growth: Evidence from China. *Journal of Development Economics*, 97(2):281–292.

Javorcik, B. (2004). Does Foreign Direct Investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review*, 94(3):605–608.

Kerr, S. P. and Kerr, W. R. (2015). Global collaborative patents. NBER Working Paper 21735. National Bureau of Economic Research Cambridge, MA.

Kerr, W. R. (2008). Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics*, 90(3):518–537.

Kuznets, S. (1960). Population change and aggregate output. In Demographic and economic change in developed countries. Columbia University Press. (p. 324-351). National Bureau of Economic Research Cambridge, MA.

Lodigiani, E. (2008). Diaspora externalities and technology diffusion. *Economie Internationale*, 115(3):43–64.

Lodigiani, E. and Salomone, S. (2015). Migration-induced transfers of norms: The case of female political empowerment. Working Papers No. 19. Department of Economics, University of Venice Ca'Foscari.

Madsen, J. B. (2007). Technology spillover through trade and TFP convergence: 135 years of evidence for the OECD countries. *Journal of International Economics*, 72:464–480.

Mayr, K. and Peri, G. (2009). Brain drain and brain return: Theory and application to Eastern-Western Europe. *The BE Journal of Economic Analysis & Policy*, 9(1):1–52.

Miguelez, E. and Moreno, R. (2015). Knowledge flows and the absorptive capacity of regions. *Research Policy*, 44(1):833–848.

Miguélez, E. (2016). Inventor diasporas and the internationalization of technology. *World Bank Economic Review.* doi: https://doi.org/10.1093/wber/lhw013

Naghavi, A. and Strozzi, C. (2015). Intellectual property rights, diasporas, and domestic innovation. *Journal of International Economics*, 96(1):150–161.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49:1471–1426.

Özden, Ç., Parsons C. R., Schiff, M., Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960-2000. *World Bank Economic Review*, 25(1):12–56.

Poncet, S. and Starosta de Waldemar, F. (2013). Export upgrading and growth: The prerequisite of domestic embeddedness. *World Development*, 51:104–118.

Ratha, D. and Shaw, W. (2007). South-South migration and remittances. World Bank Working Paper 102, World Bank, Washington, DC.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5):71–102.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, 9(1):86–136.

Santos Silva, J. M. C., and Tenreyro S. (2006). "The Log of Gravity". *Review of Economics and Statistics*, 88(4):641–658.

Simon, J. (1977). The Economics of population growth. Princeton University Press.

Spilimbergo, A. (2009). Democracy and foreign education. *The American Economic Review*, 99(1):528–543.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2):469–529.

World Bank. Global economic prospects 2008: Technology diffusion in the developing world. Washington, DC: The World Bank; 2008.

Tuccio, M. and Wahba, J. (2016). Can I have permission to leave the house? Return migration and the transfer of gender norms. IZA Discussion Paper 9216. Institute for the Study of Labor (IZA), Bonn, Germany.

Vu, K., M. (2013). A note on interpreting the Beta-convergence effect. *Economic Letters*, 118(1):46–49.

Xu, B. (2010). The sophistication of exports: Is China special? *China Economic Review*, 21(3):482–493.

Appendix to chapter 4

Table C4: Summary statistics

	Mean	Std. Dev.	Min.	Max.	N
$ECI_{i,t}$	0.020	1.047	-2.783	2.582	600
$\overline{ECI}_{i,t-5}$	0.058	0.088	0.001	0.586	600
$log(GDP_{i,t})$	8.650	1.215	5.410	11.539	600
$log(Pop_{i,t})$	2.675	1.427	-0.690	7.184	600
$Hum_{i,t}$	6.912	3.078	0.534	13.270	600
$log(Trade_{i,t})$	4.178	0.529	2.406	6.071	600
$FDI_{i,t}$	3.071	4.949	-16.071	41.065	600

Source: Author's elaboration. ECI is the Economic Complexity Index from Hausmann *et al.* (2011). $\overline{ECI}_{i,t-5}$ is the weighted-average ECI in migrants' destination countries where the weights are the emigration rates computed using Brücker *et al.* (2013). $GDP_{i,t}$ and $Pop_{i,t}$ are respectively the GDP per capita at current PPPs and the total population aged 25 years and older in millions, respectively, taken from the Penn World Table 8.0. $Hum_{i,t}$ is the average years of schooling attained for population aged 25 and over taken from Barro and Lee (2010). $Trade_{i,t}$ and $FDI_{i,t}$ are respectively the sum of exports and imports of goods and services, respectively, measured as a share of GDP and the FDI net inflows in current US dollars measured as a share of GDP, from the World Development Indicators.

Table C5: Main variables

Variable	Description	Definition and Source
$ECI_{i,t}$	Economic Complexity Index	Measure of the knowledge in a society that gets translated into the products it makes. Hausmann, Hidalgo, Bustos, Coscia, Chung, Jimenez, Simoes and Yildirim (2011).
$\overline{ECI}^{\rho}_{i,t}$	Average of the Economic Complexity Index at destination	Weighted average of the Economic Complexity Index where the weights are emigration rates ($\rho = \phi$), emigration shares ($\rho = S$), import shares ($\rho = IMP$), inverse of geographical distances ($\rho = DIS$), FDI shares ($\rho = FDI$) inverse of genetic distance ($\rho = GEN$) and initial value in 1980 ($\rho = INI$), respectively. Authors' calculations.
$M_{i,t}$	Emigration rate (log)	Proportion of migrants over the pre-migration population (25 years and older). Brücker *et al.* (2013).
$GDP_{i,t}$	GDP per Capita	Output-side real GDP per capita at current PPPs (in mil. 2005 US dollars). Penn World Table 8.0.
$Pop_{i,t}$	Total population 25 years and older	Penn World Table 8.0.
$Hum_{i,t}$	Adult Education	Educational Attainment for Population Aged 25 and Over: Average Years of Schooling Attained. Barro and Lee (2010).
$Trade_{i,t}$	Trade Openness	Sum of exports and imports of goods and services measured as a share of gross domestic product. World Development Indicators.
$FDI_{i,t}$	Foreign Direct Investment	FDI, net inflows in current U.S. dollars as a share of gross domestic product. Foreign direct investments are the net inflows of investments needed to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. World Development Indicators.
		Gravity model *à la* Feyrer (2009)
$log(Stock_{i,j,t})$	Bilateral migration stock (log)	Stock of migrants from i residing in j at time t. Özden *et al.* (2011) and Brücker *et al.* (2013).
$log(Dist_{i,j})$	Distance (log)	Geographical distance between the biggest cities of the coutries i and j weighted by the share of the city in the total population of the two countries. Head *et al.* (2010).
$Bord_{i,j}$	Common border	Dummy equal to 1 if countries i and j share a common border and 0 otherwise. Head *et al.* (2010).
$Colony_{i,j}$	Colonial past	Dummy equal to 1 if countries i and j share a colonial past and 0 otherwise. Head *et al.* (2010).
$Lang_{i,j}$	Common language	Dummy equal to 1 if at least 9% in countries i and j populations speak a common language and 0 otherwise. Head *et al.* (2010).

Table C6: Gravity model following Feyrer (2009)

	(1)
	$log(Stock_{i,j,t})$
$log(Dist_{i,j}) \times I_{1980}$	-0.844***
	(0.128)
$log(Dist_{i,j}) \times I_{1985}$	-0.759***
	(0.117)
$log(Dist_{i,j}) \times I_{1990}$	-0.741***
	(0.114)
$log(Dist_{i,j}) \times I_{1995}$	-0.738***
	(0.108)
$log(Dist_{i,j}) \times I_{2000}$	-0.729***
	(0.103)
$log(Dist_{i,j}) \times I_{2005}$	-0.689***
	(0.101)
$log(Dist_{i,j}) \times I_{2010}$	-0.690***
	(0.100)
$Bord_{i,j}$	0.357
	(0.247)
$Lang_{i,j}$	1.215***
	(0.183)
$Colony_{i,j}$	1.200***
	(0.181)
Constant	11.624***
	(1.032)
Observations	26600
Nb. origin	191
Nb. destination	20
R-squared	0.792
Year dummies	Yes
Origin dummies	Yes
Destination dummies	Yes

Source: Author's elaboration on Hausmann *et al.* (2011), Brücker *et al.* (2013) and Head *et al.* (2010). Notes: Standard errors in parentheses are clustered at the country-pair level (*** p<0.01, ** p<0.05, * p<0.1). $Dist_{i,j}$ is the great-circle distance between the capitals of i and j. I_t are year dummies. $Bord_{i,j}$ is a dummy variable equal to one if i and j share a common border. $Lang_{i,j}$ is a dummy variable equal to one if at least 9% of the populations of i and j speak a common language. $Colony_{i,j}$ is a dummy variable equal to one if i and j share a colonial past.

Table C7: Countries in the sample

Country	Observations				
Developed Countries (37)					
Australia	6	Hungary	4	Portugal	6
Austria	6	Ireland	6	Qatar	4
Belgium	2	Israel	6	Saudi Arabia	6
Canada	6	Italy	6	Slovakia	3
China, Hong Kong SAR	3	Japan	6	Slovenia	3
Croatia	3	Korea	6	Spain	6
Czech Republic	3	Kuwait	6	Sweden	6
Denmark	6	Latvia	3	Switzerland	6
Estonia	3	Netherlands	6	Trinidad and Tobago	6
Finland	6	New Zealand	6	United Kingdom	6
France	6	Norway	6	United States	6
Germany	3	Oman	6		
Greece	6	Poland	5		

Country	Observations				
Developing Countries (83)					
Albania	4	Guinea-Bissau	5	Pakistan	6
Angola	5	Honduras	6	Panama	6
Argentina	6	India	6	Paraguay	4
Azerbaijan	3	Indonesia	6	Peru	6
Bangladesh	6	Iran	5	Philippines	6
Belarus	3	Jamaica	4	Romania	5
Bolivia	6	Jordan	6	Russia	3
Bosnia and Herzegovina	3	Kazakhstan	3	Senegal	6
Botswana	1	Kenya	6	South Africa	6
Brazil	6	Kyrgyzstan	3	Sri Lanka	6
Bulgaria	5	Laos	6	Sudan	6
Cambodia	4	Lebanon	2	Syria	5
Cameroon	6	Liberia	3	Tajikistan	2
Chile	6	Lithuania	3	Tanzania	5
China	6	Macedonia	3	Thailand	6
Colombia	6	Madagascar	6	Tunisia	6
Congo, Rep. of the	6	Malawi	5	Turkey	6
Costa Rica	6	Malaysia	6	Turkmenistan	3
Cote d'Ivoire	6	Mali	5	Uganda	5
Dominican Republic	6	Mauritania	6	Ukraine	3
Ecuador	6	Mauritius	6	Uruguay	6
Egypt	6	Mexico	6	Uzbekistan	3
El Salvador	6	Moldova	3	Venezuela	6
Ethiopia	4	Mongolia	4	Vietnam	5
Gabon	6	Morocco	6	Yemen	2
Georgia	3	Mozambique	6	Zambia	6
Ghana	6	Namibia	1	Zimbabwe	6
Guatemala	6	Nigeria	6		

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: "Observations" indicates the number of observations for each country in the baseline sample in the System GMM regressions reported in col. 4 of Table 1. Countries are classified as developing countries, following the World Bank classification in 2010.

Table C8: Correlations between technological norms computed using different weights

Variables	$\overline{ECI}_{i,t-5}$	$\overline{ECI}_{i,t-5}^{S}$	$\overline{ECI}_{i,t-5}^{IMP}$	$\overline{ECI}_{i,t-5}^{FDI}$	$\overline{ECI}_{i,t-5}^{GEN}$	$\overline{ECI}_{i,t-5}^{DIS}$
$\overline{ECI}_{i,t-5}$	1.000					
$\overline{ECI}_{i,t-5}^{S}$	0.165	1.000				
	(0.000)					
$\overline{ECI}_{i,t-5}^{IMP}$	0.149	0.583	1.000			
	(0.000)	(0.000)				
$\overline{ECI}_{i,t-5}^{FDI}$	0.072	0.500	0.654	1.000		
	(0.106)	(0.000)	(0.000)			
$\overline{ECI}_{i,t-5}^{GEN}$	0.087	-0.045	0.011	0.052	1.000	
	(0.033)	(0.277)	(0.798)	(0.242)		
$\overline{ECI}_{i,t-5}^{DIS}$	0.205	-0.020	0.134	0.152	0.616	1.000
	(0.000)	(0.617)	(0.001)	(0.001)	(0.000)	

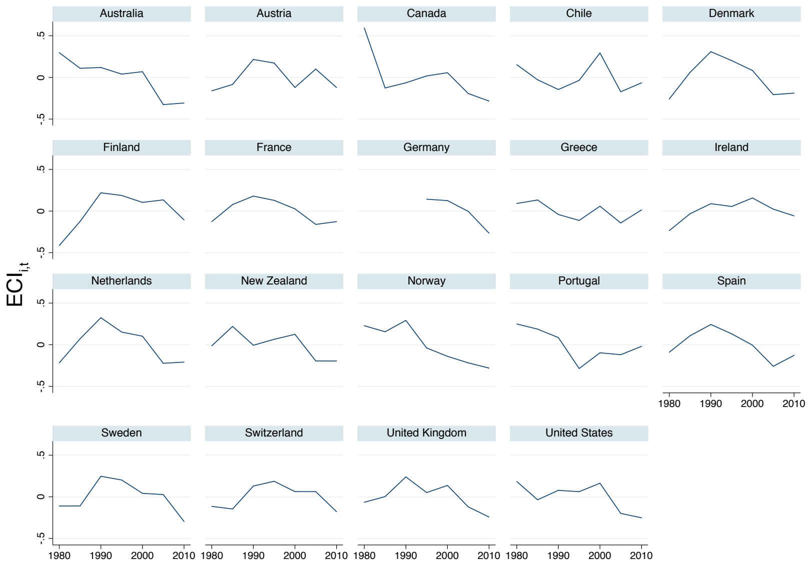
Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: pvalues in parentheses. $\overline{ECI}_{i,t-5}^{\rho}$ is the weighted-average value of ECI in migrants' destination countries where the weights are respectively emigration shares, imports, FDI and the inverse of genetic and geographical distances with $\rho = S; IMP; FDI; GEN; DIS$.

Table C9: Robustness to additional proxies for integration
System GMM with external instrument (Dep= $ECI_{i,t}$)

	(1) System GMM	(2) System GMM	(3) System GMM	(4) System GMM	(5) System GMM
$ECI_{i,t-5}$	0.756***	0.720***	0.665***	0.779***	0.695***
	(0.072)	(0.076)	(0.087)	(0.052)	(0.063)
$\overline{ECI}_{i,t-5}$	0.614**	0.831***	0.825**	0.559*	0.763***
	(0.284)	(0.269)	(0.321)	(0.289)	(0.212)
$\overline{ECI}^{IMP}_{i,t-5}$		-0.138			0.003
		(0.108)			(0.126)
$\overline{ECI}^{DIS}_{i,t-5}$			17.403***		16.817***
			(6.129)		(5.841)
$\overline{ECI}^{FDI}_{i,t-5}$				0.061	-0.028
				(0.124)	(0.127)
$log(GDP_{i,t})$	0.045	0.068	0.073	0.110***	0.110***
	(0.053)	(0.047)	(0.049)	(0.036)	(0.040)
$log(Pop_{i,t})$	0.084**	0.078**	0.102***	0.060***	0.094***
	(0.036)	(0.032)	(0.038)	(0.020)	(0.028)
$Hum_{i,t}$	0.062***	0.060**	0.047**	0.024	0.021
	(0.023)	(0.024)	(0.021)	(0.022)	(0.018)
$log(Trade_{i,t})$	-0.100	-0.096	-0.136	-0.007	-0.042
	(0.090)	(0.088)	(0.090)	(0.068)	(0.088)
$FDI_{i,t}$	0.009	0.003	0.003	-0.003	-0.009
	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)
Observations	600	600	600	406	406
Nb. countries	120	120	120	118	118
Nb. instruments	77	85	85	76	92
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000
AR(2)	0.431	0.423	0.444	0.291	0.247
Hansen J (p-value)	0.278	0.215	0.184	0.203	0.522

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level (*** p<0.01, ** p<0.05, * p<0.1). The weighted-average ECI at destination is instrumented using its predicted value obtained from a pseudo-gravity equation. The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. All the other variables are treated as endogenous and instrumented with their own second lag. FDI data had no record for Tajikistan or Turkmenistan or the year 1980.

Figure C4: Economic Complexity Index, year and country fixed effects partialled out



Source: Author's elaboration on Hausmann *et al.* (2011). Notes: The solid lines are the predicted residuals from an equation that regress the ECI on year and country fixed effects.

Table C10: System GMM with alternative lags structures (Dep= $ECI_{i,t}$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
$ECI_{i,t-5}$	0.722***	0.692***	0.693***	0.729***	0.700***	0.703***	0.733***	0.699***	0.703***	0.725***	0.695***	0.699***	0.724***	0.693***	0.698***
	(0.077)	(0.074)	(0.069)	(0.072)	(0.071)	(0.069)	(0.067)	(0.071)	(0.066)	(0.069)	(0.071)	(0.066)	(0.069)	(0.072)	(0.066)
$\overline{ECI}_{i,t-5}$	0.913*	1.008***	0.961***	0.852*	0.902**	0.903***	0.742*	0.896**	0.933***	0.920**	0.927***	0.962***	0.940**	0.940***	0.982***
	(0.471)	(0.382)	(0.358)	(0.440)	(0.365)	(0.339)	(0.445)	(0.349)	(0.339)	(0.416)	(0.333)	(0.344)	(0.416)	(0.333)	(0.350)
$log(GDP_{i,t})$	0.033	0.050	0.073	0.050	0.059	0.081*	0.054	0.074	0.082*	0.047	0.070	0.078*	0.047	0.070	0.079*
	(0.065)	(0.053)	(0.048)	(0.055)	(0.050)	(0.047)	(0.050)	(0.050)	(0.044)	(0.052)	(0.050)	(0.045)	(0.052)	(0.050)	(0.044)
$log(Pop_{i,t})$	0.091**	0.110***	0.119***	0.095***	0.118***	0.123***	0.106***	0.126***	0.128***	0.109***	0.124***	0.129***	0.108***	0.124***	0.129***
	(0.035)	(0.029)	(0.029)	(0.034)	(0.030)	(0.029)	(0.034)	(0.031)	(0.030)	(0.035)	(0.031)	(0.031)	(0.035)	(0.030)	(0.030)
$Hum_{i,t}$	0.074***	0.077***	0.066***	0.065***	0.072***	0.062***	0.063**	0.066***	0.061***	0.066***	0.068***	0.063***	0.066***	0.068***	0.062***
	(0.026)	(0.023)	(0.021)	(0.025)	(0.022)	(0.021)	(0.024)	(0.021)	(0.021)	(0.024)	(0.022)	(0.021)	(0.024)	(0.022)	(0.021)
$log(Trade_{i,t})$	-0.130	-0.081	-0.040	-0.103	-0.050	-0.017	-0.081	-0.034	-0.017	-0.091	-0.043	-0.018	-0.092	-0.046	-0.021
	(0.103)	(0.083)	(0.072)	(0.087)	(0.078)	(0.068)	(0.081)	(0.076)	(0.069)	(0.083)	(0.077)	(0.070)	(0.083)	(0.077)	(0.071)
$FDI_{i,t}$	0.011*	0.008	0.003	0.010	0.007	0.002	0.009	0.007	0.003	0.010	0.006	0.003	0.010	0.006	0.003
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
Nb. Lags $ECI_{i,t-5}$	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Nb. Lags Endogenous Var.	2	3	4	2	3	4	2	3	4	2	3	4	2	3	4
Observations	600	600	600	600	600	600	600	600	600	600	600	600	600	600	600
Nb. countries	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
Nb. instruments	74	97	114	78	101	118	81	104	121	83	106	123	84	107	124
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.433	0.448	0.463	0.445	0.467	0.469	0.454	0.472	0.470	0.450	0.466	0.472	0.449	0.463	0.468
Hansen J (p-value)	0.201	0.171	0.263	0.175	0.219	0.272	0.170	0.213	0.335	0.213	0.264	0.394	0.237	0.289	0.431

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level (*** $p<0.01$, ** $p<0.05$, * $p<0.1$). The lagged dependent variable is treated as predetermined. All other variables are treated as endogenous and instrumented using their own second lag.

Table C11: Control variables added one by one

	(1) System GMM ECI	(2) System GMM ECI	(3) System GMM ECI	(4) System GMM ECI	(5) System GMM ECI	(6) System GMM ECI	(7) System GMM ECI	(8) System GMM ECI	(9) System GMM ECI
$ECI_{i,t-5}$	0.724***	0.929***	0.200***	0.871***	0.908***	0.757***	0.729***	0.906***	0.861***
	(0.069)	(0.015)	(0.059)	(0.048)	(0.027)	(0.063)	(0.079)	(0.054)	(0.043)
$\overline{ECI}_{i,t-5}$	0.940**	0.295**	2.241***	0.628*	0.869*	0.510	0.707	0.392	0.251
	(0.416)	(0.132)	(0.736)	(0.343)	(0.520)	(0.390)	(0.464)	(0.460)	(0.562)
$log(Pop_{i,t})$	0.108***				0.033				
	(0.035)				(0.060)				
$log(GDP_{i,t})$	0.047					0.176***			
	(0.052)					(0.057)			
$Hum_{i,t}$	0.066***						0.093***		
	(0.024)						(0.034)		
$log(Trade_{i,t})$	-0.092							-0.065	
	(0.083)							(0.162)	
$FDI_{i,t}$	0.010								0.018*
	(0.006)								(0.011)
Observations	600	600	600	600	600	600	600	600	600
Nb. countries	120	120	120	120	120	120	120	120	120
Nb. instruments	84			34	44	44	44	44	44
R-squared		0.893	0.084						
Country fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000			0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.449			0.305	0.287	0.388	0.324	0.448	0.393
Hansen J (p-value)	0.237			0.004	0.023	0.015	0.007	0.015	0.011

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level. (*** p<0.01, ** p<0.05, * p<0.1). The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags.

Table C12: Control variables added two by two

	(1) System GMM ECI	(2) System GMM ECI	(3) System GMM ECI	(4) System GMM ECI	(5) System GMM ECI	(6) System GMM ECI	(7) System GMM ECI	(8) System GMM ECI	(9) System GMM ECI	(10) System GMM ECI
$ECI_{i,t-5}$	0.795***	0.621***	0.942***	0.914***	0.665***	0.813***	0.808***	0.805***	0.775***	0.903***
	(0.053)	(0.078)	(0.028)	(0.028)	(0.085)	(0.059)	(0.067)	(0.064)	(0.068)	(0.038)
$\overline{ECI}_{i,t-5}$	1.037**	1.606***	0.483	0.433	0.873*	0.299	0.169	0.186	0.320	0.442
	(0.504)	(0.526)	(0.424)	(0.544)	(0.493)	(0.407)	(0.388)	(0.440)	(0.557)	(0.511)
$log(Pop_{i,t})$	0.052	0.185***	0.029	0.012						
	(0.039)	(0.055)	(0.035)	(0.049)						
$log(GDP_{i,t})$	0.137***				0.182**	0.118**	0.116*			
	(0.050)				(0.073)	(0.049)	(0.068)			
$Hum_{i,t}$		0.109***			0.059			0.062**	0.060**	
		(0.027)			(0.037)			(0.025)	(0.025)	
$log(Trade_{i,t})$			-0.078			-0.063		-0.209		-0.111
			(0.135)			(0.103)		(0.131)		(0.122)
$FDI_{i,t}$				0.018*			-0.001		0.014	0.022**
				(0.010)			(0.010)		(0.008)	(0.009)
Observations	600	600	600	600	600	600	600	600	600	600
Nb. countries	120	120	120	120	120	120	120	120	120	120
Nb. instruments	54	54	54	54	54	54	54	54	54	54
Country fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.359	0.353	0.420	0.374	0.442	0.500	0.337	0.366	0.385	0.510
Hansen J (p-value)	0.060	0.091	0.076	0.029	0.020	0.031	0.085	0.0152	0.006	0.132

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level. (*** p<0.01, ** p<0.05, * p<0.1). The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. Independent variables are treated as endogenous variables and instrumented using their own second lag.

Table C13: Control variables added three by three

	(1) System GMM ECI	(2) System GMM ECI	(3) System GMM ECI	(4) System GMM ECI	(5) System GMM ECI	(6) System GMM ECI	(7) System GMM ECI	(8) System GMM ECI	(9) System GMM ECI	(10) System GMM ECI
$ECI_{i,t-5}$	0.600***	0.834***	0.870***	0.742***	0.763***	0.922***	0.763***	0.744***	0.809***	0.802***
	(0.073)	(0.056)	(0.051)	(0.080)	(0.071)	(0.026)	(0.061)	(0.078)	(0.067)	(0.054)
$\overline{ECI}_{i,t-5}$	1.693***	0.748**	0.845*	0.558	0.924*	0.736*	0.373	0.332	0.091	0.297
	(0.489)	(0.368)	(0.437)	(0.492)	(0.490)	(0.398)	(0.392)	(0.404)	(0.378)	(0.505)
$log(Pop_{i,t})$	0.160***	0.059**	0.064*	0.099**	0.088*	0.054				
	(0.047)	(0.029)	(0.038)	(0.046)	(0.049)	(0.042)				
$log(GDP_{i,t})$	0.068	0.112**	0.052				0.085	0.096	0.112*	
	(0.068)	(0.044)	(0.058)				(0.055)	(0.073)	(0.060)	
$Hum_{i,t}$	0.096***			0.080***	0.060**		0.049*	0.051		0.072***
	(0.029)			(0.027)	(0.024)		(0.027)	(0.033)		(0.024)
$log(Trade_{i,t})$		-0.003		-0.141		-0.102	-0.205*		-0.069	-0.227**
		(0.115)		(0.114)		(0.113)	(0.114)		(0.086)	(0.110)
$FDI_{i,t}$			0.010		0.012	0.021**		-0.003	0.006	0.019***
			(0.007)		(0.008)	(0.009)		(0.010)	(0.009)	(0.007)
Observations	600	600	600	600	600	600	600	600	600	600
Nb. countries	120	120	120	120	120	120	120	120	120	120
Nb. instruments	64	64	64	64	64	64	64	64	64	64
Country fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.392	0.517	0.359	0.406	0.380	0.489	0.414	0.339	0.458	0.409
Hansen J (p-value)	0.207	0.096	0.219	0.080	0.025	0.387	0.040	0.079	0.191	0.090

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level. (*** p<0.01, ** p<0.05, * p<0.1). The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. Independent variables are treated as endogenous variables and instrumented using their own second lag.

Table C14: Control variables added four by four

	(1) System GMM ECI	(2) System GMM ECI	(3) System GMM ECI	(4) System GMM ECI	(5) System GMM ECI
$ECI_{i,t-5}$	0.691***	0.744***	0.848***	0.751***	0.766***
	(0.082)	(0.072)	(0.057)	(0.064)	(0.068)
$\overline{ECI}_{i,t-5}$	0.931**	1.152***	0.736*	0.766*	0.318
	(0.430)	(0.437)	(0.381)	(0.441)	(0.411)
$log(Pop_{i,t})$	0.119***	0.111***	0.063**	0.107***	
	(0.043)	(0.038)	(0.031)	(0.032)	
$log(GDP_{i,t})$	0.064	0.017	0.091*		0.058
	(0.055)	(0.059)	(0.050)		(0.057)
$Hum_{i,t}$	0.075***	0.062**		0.073***	0.060**
	(0.028)	(0.024)		(0.022)	(0.025)
$log(Trade_{i,t})$	-0.121		-0.069	-0.104	-0.205**
	(0.113)		(0.094)	(0.085)	(0.089)
$FDI_{i,t}$		0.009	0.011	0.012*	0.011
		(0.007)	(0.007)	(0.006)	(0.008)
Observations	600	600	600	600	600
Nb. countries	120	120	120	120	120
Nb. instruments	74	74	74	74	74
Country fixed-effects	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000
AR(2)	0.454	0.374	0.467	0.434	0.399
Hansen J (p-value)	0.140	0.190	0.250	0.194	0.192

Source: Author's elaboration on Hausmann *et al.* (2011) and Brücker *et al.* (2013). Notes: Standard errors in parentheses are clustered at the country level. (*** p<0.01, ** p<0.05, * p<0.1). The lagged dependent variable is treated as predetermined and instrumented with its own first to five lags. Independent variables are treated as endogenous variables and instrumented using their own second lag.

Chapter 5

Conclusions

À l'heure où la population des pays en développement ne cesse d'augmenter, où les inégalités entre pays sont toujours aussi importantes, et où, les changements climatiques à venir seront probablement responsables de déplacements jamais encore observés, le sujet des migrations internationales n'a jamais paru autant d'actualité. D'aucuns s'accordent ainsi à dire que, malgré des politiques toujours plus restrictives, les pays de l'OCDE feront face, dans les prochaines décennies, à une pression migratoire sans précédent.

Si la recherche en économie a produit, depuis longtemps déjà, un nombre important de résultats sur le sujet des migrations internationales, de nombreuses questions restent encore en suspens quant à l'impact des migrations internationales sur les migrants eux-mêmes, leur pays d'origine et les natifs de leur pays d'accueil. La présente thèse vient donc contribuer à cet effort au travers des trois chapitres empiriques précédemment développés.

Le chapitre 2 de cette thèse se concentre sur les effets du multiculturalisme sur la croissance économique. Nous exploitons les variations observées dans l'évolution de la diversité culturelle aux États-Unis pour la période 1960-2010, entre les différents États. Nos résultats ne mettent pas en évidence d'effet économique majeur de la diversité sur la croissance économique. Si l'effet est bel et bien significatif et positif, mais de faible ampleur, pour les migrants les plus éduqués, ce dernier n'est pas significatif pour les migrants n'ayant pas de diplôme du tertiaire. De plus, nos résultats réfutent l'existence d'effets de contamination par la migration. Malgré l'utilisation de données en longue période et une meilleure prise en compte de l'hétérogénéité inobservée, nos résultats corroborent donc la littérature économique existante. Or, si l'impact positif de la diversité chez les diplômés du tertiaire est alors un résultat établi, trop peu de papiers s'attachent encore à mettre en lumière les canaux aux travers desquels la diversité impacte la croissance économique des pays d'accueil des migrants. De futures recherches, utilisant notamment des données micro-économiques, au niveau des firmes par exemple, devront pouvoir combler ce manque. D'une manière générale, si l'impact macro-économique des migrations internationales sur les économies des pays d'accueil semble faible, de futurs travaux devront pouvoir réconcilier les effets non significatifs de la recherche avec les perceptions négatives que semblent exprimer les natifs envers l'immigration. Concernant des recommandations de politiques économiques possibles, les résultats de ce chapitre soulignent l'importance de la prise en compte du pays d'origine des migrants les plus éduqués dans les politiques migratoires des pays de l'OCDE. Aux États-Unis le programme « Diversity Visa » créé suite à « l'Immigration Act » de 1990 répond à cela en proposant un canal de migration privilégié pour les groupes sous-représentés dans la population Américaine. Ainsi, chaque année, environ 55 000 visas sont aléatoirement alloués à des pays ayant envoyé moins de 50 000 migrants aux États-Unis sur les 5 dernières années. Cependant, ce programme ne rejoint pas entièrement nos résultats puisque ces visas sont réservés à des individus ayant au moins une éducation secondaire et non seulement aux seuls diplômés du tertiaire.

Le Chapitre 3 se focalise sur les attitudes que les natifs expriment envers l'immigration et les migrants eux-mêmes. En effet, alors que la littérature économique a largement étudié l'effet de la discrimination sur l'intégration sur le marché du travail, peu de papiers se sont concentrés sur l'hétérogénéité observée dans les niveaux de discrimination auxquels font face les migrants. Notre étude fait exception à cette règle en utilisant les variations dans la confiance qu'attribuent les Allemands envers les différents pays d'origine des migrants et les variations dans les niveaux de confiance au sein d'une même origine dans différentes régions. A l'aide d'une analyse de survie au niveau mensuel sur la période 1984-2012, nos résultats montrent un effet significatif et négatif de la confiance sur les durées de chômage des migrants en Allemagne. Concernant la magnitude de l'effet, si les Allemands avaient autant confiance envers les Turcs qu'envers les immigrés Autrichiens alors les Turcs verraient leurs durées de chômage réduites de trois mois. L'attitude des natifs envers les immigrés conditionne donc leur intégration sur le marché du travail, et par conséquent, leur impact sur l'économie toute entière. Ces attitudes agissent notamment de manière perverse en réduisant les incitations à l'investissement en capital humain à destination, principal déterminant de l'assimilation économique des migrants. De futures recherches devront alors évaluer les coûts directs et indirects de ces attitudes pour les pays receveurs de migrants. En effet, la littérature économique sur les discriminations n'aborde que marginalement ce sujet alors que celui-ci est central pour évaluer l'importance des politiques d'intégration. Plusieurs recommandations de politiques peuvent néanmoins être faites sur la base de ces premiers résultats. En fonction du type de discrimination auquel font face les migrants (de goût vs statistique) des mesures comme les CVs anonymes ou encore les équivalences de diplômes entre pays peuvent atténuer les effets de la discrimination sur l'intégration sur le marché du travail. L'Allemagne se place justement comme un leader en Europe sur ce plan à travers les engagements pris en 2012 dans le « Recognition Act » qui permet aux travailleurs étrangers de mieux valoriser leurs compétences professionnelles et leurs diplômes sur le marché du travail. Cependant, beaucoup reste encore à faire, et l'index MIPEX (Migration Integration Policy Index) pour l'année 2015, qui note 38 pays sur la base de 167 indicateurs différents sur les efforts faits pour l'intégration des migrants, classe l'Allemagne 22$^{\text{ème}}$ sur le volet des mesures anti-discriminations, loin derrière des pays comme le Canada ou les USA. D'autre part, une étude récente dans le contexte du Japon par Facchini Giovanni *et al.* (2016)[1] montre que des campagnes d'information sur les bénéfices potentiels de la migration peuvent aussi être à même de réduire les attitudes négatives exprimées par les natifs envers l'immigration.

Le Chapitre 4 enfin, se concentre sur les retombées de la migration attendues dans les pays d'origine des migrants. Parmi ces retombées, la littérature économique s'est notamment intéressée à l'impact des migrations internationales sur l'évolution de la technologie dans les pays d'origine des migrants. Ce dernier chapitre s'inscrit dans ce courant en proposant une analyse macroéconomique au niveau pays sur la période 1980-2010. En effet, cette période a été marquée, à la fois par une augmentation de l'immigration des pays en

[1] Facchini G., Margalit Y., et Nakata H. (2016). Countering Public Opposition to Immigration: The impact of Information Campaigns. *IZA Discussion paper No. 10420.*

développement vers les pays de l'OCDE, à fort potentiel technologique, mais aussi par une convergence dans les niveaux de technologies observés au niveau mondial. Nos résultats font le lien entre ces deux faits stylisés en montrant comment, pour un pays donné, l'augmentation de sa diaspora dans les 20 pays de l'OCDE est associée à une augmentation de son stock de connaissances productives. Nos résultats mettent aussi en avant le fait que les transferts technologiques sont plus à même d'exister lorsque les taux d'émigration sont forts et que la technologie à destination est importante. Une fois n'est pas coutume, si la littérature reconnait le rôle des migrants dans l'évolution de la technologie dans leur pays d'origine, les mécanismes de ces transferts sont encore mal connus. De futures recherches devront alors chercher à mettre en évidence les canaux par lesquels les migrants améliorent les connaissances productives dans leur pays d'origine. Les migrations retours, la baisse des coûts de communication, l'augmentation des échanges ou encore les partenariats industriels et scientifiques apparaissent alors comme des candidats naturels. Un raisonnement trop rapide consisterait à recommander aux pays en développement de faciliter de manière importante l'émigration vers des pays avec de forts niveaux technologiques. Or, rien n'assure que les transferts technologiques de la diaspora compensent entièrement les pertes directes en capital humain que peuvent engendrer les départs des individus les plus éduqués des pays en développement vers les pays de l'OCDE. De futures recherches devront conclure quant à un effet positif ou négatif net de l'émigration en général sur la technologie des pays en développement. Une recommandation de politique économique plus prudente serait alors d'encourager les pays en développement à intensifier les liens avec leurs diasporas déjà installées à l'étranger.

Pour conclure, les trois essais de cette thèse, s'ils traitent de sujets qui peuvent parfois paraitre éloignés, ont tous en commun, non seulement le thème des migrations internationales, mais aussi et surtout l'idée centrale, comme le souligne G. Borjas dans son livre « We Wanted Workers », que les immigrés ne sont pas seulement des travailleurs mais bien des individus à part entière. Leur culture, leurs normes, la diversité qu'ils apportent à leur pays d'accueil, ainsi que leurs interactions avec les natifs et leurs proches restés dans leur pays d'origine, ne peuvent alors être ignorés. Il en va d'une évaluation juste des effets globaux de la migration au niveau international.